



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C07H 21/02, 21/04, C12N 5/00, 5/04, 5/06, 5/10, 5/16, 15/00, 15/09, 15/10, 15/11, 15/12, C12P 21/04, 21/06		A1	(11) International Publication Number: WO 98/56804
			(43) International Publication Date: 17 December 1998 (17.12.98)
(21) International Application Number: PCT/US98/12125		(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).	
(22) International Filing Date: 11 June 1998 (11.06.98)		(72) Inventors; and	
(30) Priority Data:		(75) Inventors/Applicants (for US only): MOORE, Paul, A. [GB/US]; Apartment 104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mt. Nebo Road, Poolesville, MD 20837 (US). YOUNG, Paul [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13202 L Astoria Hill Court, Germantown, MD 20874 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Damestown, MD 20878 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US).	
60/049,547	13 June 1997 (13.06.97)	US	(74) Agents: BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).
60/049,548	13 June 1997 (13.06.97)	US	
60/049,549	13 June 1997 (13.06.97)	US	
60/049,550	13 June 1997 (13.06.97)	US	
60/050,566	13 June 1997 (13.06.97)	US	
60/049,606	13 June 1997 (13.06.97)	US	
60/049,607	13 June 1997 (13.06.97)	US	
60/049,608	13 June 1997 (13.06.97)	US	
60/049,609	13 June 1997 (13.06.97)	US	
60/049,610	13 June 1997 (13.06.97)	US	
60/049,611	13 June 1997 (13.06.97)	US	
60/050,901	13 June 1997 (13.06.97)	US	
60/052,989	13 June 1997 (13.06.97)	US	
60/051,919	8 July 1997 (08.07.97)	US	
60/055,984	18 August 1997 (18.08.97)	US	
60/058,665	12 September 1997 (12.09.97)	US	
60/058,668	12 September 1997 (12.09.97)	US	
60/058,669	12 September 1997 (12.09.97)	US	
60/058,750	12 September 1997 (12.09.97)	US	
60/058,971	12 September 1997 (12.09.97)	US	
60/058,972	12 September 1997 (12.09.97)	US	
60/058,975	12 September 1997 (12.09.97)	US	
60/060,834	2 October 1997 (02.10.97)	US	
60/060,841	2 October 1997 (02.10.97)	US	
60/060,844	2 October 1997 (02.10.97)	US	
60/060,865	2 October 1997 (02.10.97)	US	
60/061,059	2 October 1997 (02.10.97)	US	
60/061,060	2 October 1997 (02.10.97)	US	
(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).			
<p>Published With international search report.</p>			
(54) Title: 86 HUMAN SECRETED PROTEINS			
(57) Abstract			
<p>The present invention relates to 86 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

86 Human Secreted Proteins

Field of the Invention

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and 5 their production.

Background of the Invention

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or 10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of 15 proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

20 Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or 25 secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes 30 encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying 35 and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, 5 and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

10

Detailed Description

Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

15

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

20

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

25

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

30

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, 5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained 10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the 15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages 20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH₂PO₄; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even 25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include 30 Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such 35 as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5 The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and 10 double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability 15 or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

20 The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, 25 as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be 30 branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a 35 nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins

5 such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., *Meth Enzymol* 182:626-646 (1990); 10 Rattan et al., *Ann NY Acad Sci* 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

15 "A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present 20 invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

25 **Polynucleotides and Polypeptides of the Invention**

FEATURES OF PROTEIN ENCODED BY GENE NO: 1

The translation product of this gene shares sequence homology with LIM-homeobox domain proteins, such as T-cell translocation protein, which are thought to 30 be important in development and leukemogenesis. In addition, translation product of this gene shares homology with the human breast tumor autoantigen (See Accession No. gil1914877). In one embodiment the polypeptides of the invention comprise the sequence:

35 MNGSHKDPLLPFPASARTPSLPPAPPAQAPLPWKPSGFARISPPPPLAILQYRG
KADHGESGQQQLAAAPGDGRLPLLEAVRRLRGQDCGPLSALCHGQLLAQPV
VLLLPGAXGDIGTSCYTKSGMILCRNDYIRLFGNSGACSACGQSIPASELVMRA
QGNVYHLKCFTCSTCRNRLVPGDRFHYINGSLFCEHDRPTALINGHLNSLQSN

PLLPDQKVCKVRVMQNACLHLRFVHHRWIPCXFSRQVTFVASTSASSMPLHLL
(SEQ ID NO:211); MARTRTPSSPFLLRELPPSLQLRQPRRPFGSRAASLAFHRR
RLSQYCNIGEKQTMVNPGSSSQPPPVTAGSLSWKRCAGCGKIADRFLLYA
(SEQ ID NO:212); LFGNSGACSACGQSIPASELVMRA (SEQ ID NO:213);
5 HDRPTALINGHLNSLQSNP (SEQ ID NO:214); and/or LVPGDRFHYING (SEQ ID
NO:215). Polynucleotide fragments encoding these polypeptide fragments are also
encompassed by the invention.

This gene is expressed primarily in fetal brain, osteosarcoma, IL-1/TNF treated
synovial, and estradiol treated endometrial stromal cells, and to a lesser extent in
10 chondrosarcoma, smooth muscle and number of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, developmental defects or leukemia. Similarly, polypeptides and
15 antibodies directed to these polypeptides are useful in providing immunological probes
for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the hematopoietic system and immune
system, expression of this gene at significantly higher or lower levels may be routinely
detected in certain tissues and cell types (e.g., brain and other tissue of the nervous
20 system, bone cells, synovial tissue, endometrial tissue and other reproductive tissue,
cartilage cells, smooth muscle, and blood cells and cells and tissue of the immune
system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,
urine, synovial fluid or spinal fluid) or another tissue or cell sample or another tissue or
cell sample or another tissue or cell sample taken from an individual having such a
25 disorder, relative to the standard gene expression level, i.e., the expression level in
healthy tissue or bodily fluid or bodily fluid or bodily fluid from an individual not
having the disorder. Preferred epitopes include those comprising a sequence shown in
SEQ ID NO. 111 as residues: Met-1 to Cys-9.

The tissue distribution and homology to the LIM-homeodomain containing
30 proteins, such as T-cell translocation factor, indicates that polynucleotides and
polypeptides corresponding to this gene are useful for diagnosis and intervention of
leukemia and other developmental defects. Because of the importance of the LIM-
homeodomain proteins in development and their correlation to number of leukemic
diseases, the molecule can be either used as a diagnostic or prognostic indicator for
35 leukemia progression or a therapeutic target. In addition, polynucleotides and
polypeptides corresponding to this gene are useful for the detection/treatment of
neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease,

Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, and autism. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or 5 disorders of the cardiovascular system. Furthermore, homology to the breast auto-antigen may suggest this gene is useful in the detection, prevention, and or treatment of breast cancer and/or other proliferative disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 2

10 Translation product of gene has homology to a highly conserved member of the human calpain family of proteases, Calpain large subunit 1 gene (See Accession No.T32454). Calpains are thought to play a defining role in protein regulation, particularly during development. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

15 MKYMGGCAKVMCKYYVILYQGLEYPLLXSGDPETSPPWILRADCVLSSRNFH
SNXGRLTINKIYVIGGGKYRGEVTNGAK (SEQ ID NO:216);
MGQSELYSSILRNILGVLFLVYTRGGFLLSPLLHGTLTCAHS (SEQ ID NO:217);
MVLLLLTVAASYTVFWMIGDVLDILFLWNFEYTTLY (SEQ ID NO:218);
MELYNSLCPICYFSTVLTITYYIYFVYSQSSXIRMKVP (SEQ ID NO:219);
20 MQIVIVLYCVRNKKVCTCSVQTQFFFPIFPILGCLNGCRTQE (SEQ ID NO:220); MKYMGGCAKVMCKYYVILYQGLEYPLLX (SEQ ID NO:221);
LEYPLLXSGDPET SPPWILRADCVLSSRNFHNSX (SEQ ID NO:222); and/or
RNFHSNXGRLTINKIY VIGGGKYRGEVTNGAK (SEQ ID NO:223). An
additional embodiment is the polynucleotide fragments encoding these polypeptide 25 fragments.

This gene is expressed primarily in caudate nucleus, dermatofibrosarcoma protuberance and apoptotic T-cells, and to a lesser extent in eosinophils, brain and smooth muscle.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative diseases or immune disorders. Similarly, 35 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system or immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., skin, T-cells and other blood

cells and cells and tissue of the immune system, brain and other tissue of the nervous system, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in caudate nucleus and apoptic T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection or intervention of neurodegenerative diseases and behavioral disorders such as

5 Alzheimer's Disease, Parkinson's Disease, Huntington's disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder or immune disorders, because the elevated level of the molecule in cells undergoing cell death may be the cause or consequence of these degenerative conditions. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

10 disorders associated with the developing embryo, or disorders of the cardiovascular system.

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene maps to chromosome 15, and therefore, may be used as a marker in

20 linkage analysis for chromosome 15. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: VTNEMSQGRGKYDFY IGLGLAMSSSFIGGSFILKKGLLRLARKGSMRAGQGGHA YLKEWLWWAGL LSMGAGEVANFAAYAFAPATLVTPLGALSVLVSAILSSYFLNERLNHGKIGCL LSILG STVMVIHAPKEEEIETLNE (SEQ ID NO:224);

25 VTNEMSQGRGKYDFYIGLGLAMSSSFIGGSFILKKGLLRLARKGSMRAGQGQ GHAYLKEWLWWAGL LSMGAGEVANF (SEQ ID NO:225);

NFAAYAFAPATLVTPLGALSVLVSAILSSY (SEQ ID NO:226); and/or

30 ERLNLHGKIGCLLSILGSTVMVIHAPKEEEIETLNE (SEQ ID NO:227). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments

This gene is expressed primarily in colon carcinoma cell line, and to a lesser extent in aorta endothelial cells, T-cells, human erythroleukemia cells (HEL), and stromal cells (TF274).

Therefore, polynucleotides and polypeptides of the invention are useful as

35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, colon carcinoma. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of colon carcinoma tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, aorta and other vascular tissue, T-cells and other cells and tissue of the immune system, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 113 as residues: Asn-191 to Ser-196, Asn-208 to Gly-214.

The tissue distribution in colon carcinoma indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and intervention of colon carcinoma and/or other tumors. Additionally the significant presence in T-cell populations may indicate the involvement of the function of the gene product in cancer immuno surveillance. Furthermore, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders, in general. The expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene is expressed primarily in ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive or endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive or endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 114 as residues:

5 Pro-20 to Ser-25.

The tissue distribution in ovary indicates that polynucleotides and polypeptides corresponding to this gene are useful for assessing reproductive dysfunction or endocrine disorders, because factors secreted by ovary may be involved in reproductive processes, and in cases have global hormonal effects.

10

FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in tissues in the central nervous system, including pineal gland, frontal cortex, and dura mater, and to a lesser extent in bladder, lung, T-cells and liver.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative diseases, endocrine disorders, and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are 20 useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., tissue of the nervous system, bladder, lung, liver, and T-cells and other cells and tissues of the 25 immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID 30 NO. 115 as residues: Glu-14 to Arg-20.

The primary tissue distribution in the central nerve system indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and intervention of neurodegenerative diseases or endocrinedisorders, because extracellular proteins in these tissues may function as a neurotrophic factor, a matrix 35 protein for tissue integrity, a neuroguidance factor or as a hormone.

FEATURES OF PROTEIN ENCODED BY GENE NO: 6

This gene is expressed primarily in spleen, resting T-cells, colorectal tumor and pancreatic carcinoma, and to a lesser extent in number of tissues including prostate, synovial hypoxia, osteosarcoma, ulcerative colitis, myeloid progenitor cells, lung and placenta.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammation, immunosurveillance of cancers, and immune and
10 gastrointestinal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in carcinogenesis or the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell
15 types (e.g., prostate, synovial tissue, bone cells, colon, myeloid progenitor cells, lung, cells and tissue of the immune system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an
20 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 116 as residues: Arg-29 to Pro-37, Gln-46 to Val-56.

25 The primary tissue distribution in lymphatic tissues such as T-cells and spleen, as well as tumors and ulcerative tissues indicates that the protein product of this gene may be involved in the immuno response to or immunosurveillance of carcinogenesis and/or inflammatory conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The translation product of this gene shares very weak sequence homology with voltage dependent sodium channel protein and Bowman-Birk proteinasse inhibitor
30 which is thought to be important in membrane signaling or extracellular signaling cascades. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: RFKTLMTNKSEQDGDSSKTIEISDMKYHIFQ (SEQ ID NO:228); and/or LVEGKLFYAHKVLLVTXSNR (SEQ ID NO:229) (See Accession No. gnl|PID|d1020763 (AB000216)). An additional embodiment is the
35 polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in prostate cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostate cancer. Similarly, polypeptides and antibodies directed to these 5 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of prostate cancer tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, 10 synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 117 as residues: Glu-30 to Ser-35.

15 The tissue distribution in the prostate cancer and homology to sodium channel or proteinase inhibitor suggest that polynucleotides and polypeptides corresponding to this gene are useful for the intervention of cancer progression, because the gene product may be involved in multidrug resistance by altering the drug kinetics by serving the function as a channel transporter. Alternatively, the proteinase inhibitor like function 20 may facilitate tumor metastasis. By targeting these functions, either through vaccine or small molecules, therapeutics may be rationally designed to slow the cancer progression.

FEATURES OF PROTEIN ENCODED BY GENE NO: 8

25 This gene is expressed primarily in ovary and to a lesser extent in the adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 30 not limited to, female infertility and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system and the endocrine system, expression of this gene at significantly higher or lower levels may be 35 routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene in ovary and adrenal gland indicates that

5 polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of female infertility, endocrine disorders, ovarian function, amenorrhea, ovarian cancer and metabolic disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 9

10 This gene is expressed only in prostate cancer. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostate disorders including cancer. Similarly, polypeptides and

15 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and male reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and cancerous and wounded

20 tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene only in prostate cancerous tissue, indicates

25 that polynucleotides and polypeptides corresponding to this gene are useful for the treatment/diagnosis of male infertility, metabolic disorders, and prostate disorders including benign prostate hyperplasia and prostate cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 10

30 This gene is expressed primarily in placenta and to a lesser extent in ovary. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, female infertility, pregnancy disorders, and ovarian cancer. Similarly,

35 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive

system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, and ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from 5 an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 120 as residues: Gln-39 to Gly-73.

The tissue distribution of this gene in placenta and ovary indicates that 10 polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of female infertility, endocrine disorders, fetal deficiencies, ovarian failure, amenorrhea, and ovarian cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 11

15 Gene shares homology with the gene for the Human 3' apolipoprotein B SAR element gene Rh32 (See Accession No. T31530).

This gene is expressed primarily in prostate and in the pancreas. Therefore, polynucleotides and polypeptides of the invention are useful as 20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostate and pancreatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene 25 at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and pancreas, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an 30 individual not having the disorder.

The tissue distribution of this gene in prostate and pancreas, indicates that 35 polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of male infertility, prostate disorders including benign prostate hyperplasia, prostate cancer, pancreatic cancer, type I and type II diabetes and hypoglycemia. Homology to a known human apolipoprotein may suggest this gene is useful for the detection, prevention, or treatment of various metabolic disorders,

particularly those secondary to lipoprotein disorders such as atherosclerosis, coronary heart disease, stroke, and hyperlipidemias.

FEATURES OF PROTEIN ENCODED BY GENE NO: 12

5 Gene has homology to conserved Beta-casein, an abundant milk protein (See Accession No.Q37894).

This gene is expressed primarily in stomach.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

10 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the digestive tract and/or mammary glands. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system

15 and breast, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and stomach and other gastrointestinal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

20 expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene indicates a role in the treatment/diagnosis of digestive disorders including stomach cancer and ulceration. Furthermore, the homology to conserved beta-casein may indicate this gene as having utility in the

25 diagnosis and prevention of mammary gland disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed in brain and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as

30 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disease states, behavioral abnormalities and pulmonary disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification

35 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, nervous, and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell

types (e.g., brain and other tissue of the nervous system, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive 10 compulsive disorder and panic disorder. In addition it could be used in the detection and treatment of pulmonary disease states such as lung lymphoma or sarcoma formation, pulmonary edema and embolism, bronchitis and cystic fibrosis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 14

15 This gene is expressed exclusively in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to 20 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and cancerous and 25 wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of immune disorders such as arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative 35 disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 15

This gene is expressed primarily in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 125 as residues: Ala-46 to Asp-51.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 16

This gene is expressed primarily in endometrial tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly endometrial. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrial cells and other reproductive cells or tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of ovarian and 5 other endometrial cancers, as well as reproductive dysfunction, prenatal disorders or fetal deficiencies.

FEATURES OF PROTEIN ENCODED BY GENE NO: 17

This gene is expressed primarily in a variety of osteoclastic cells: osteoclastoma 10 stromal cells, osteosarcoma, chondrosarcoma and stromal cell culture. To a lesser extent, it is also seen in a variety of fetal and embryonic cell and tissue types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 15 not limited to, bone cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and developmental systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell 20 types (e.g., bone cells, cartilage, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those 25 comprising a sequence shown in SEQ ID NO. 127 as residues: Gln-34 to Gln-41, Asn-76 to Lys-82, Ser-85 to Lys-91.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and detection of a variety disorders and conditions affecting bone and the skeletal system, including: osteoporosis, fracture, 30 osteosarcoma, osteoclastoma, chondrosarcoma, ossification and osteonecrosis, arthritis, tendonitis, chondromalacia and inflammation.

FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in smooth muscle. 35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, cardiovascular disorders including lymphatic system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the 5 cardiovascular and lymphatic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscles, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., 10 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of conditions and pathologies of the cardiovascular system: heart disease, restenosis, atherosclerosis, 15 stroke, angina, thrombosis, and wound healing.

FEATURES OF PROTEIN ENCODED BY GENE NO: 19

The translation product of this gene shares sequence homology with 5'-nucleotidase (See Accession No. 2668557) as well as the gene for alpha-1 collagen type 20 X (See Accession No. gblX67348IMMCOL10A). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence:

MAQHFSLAACDVVGFDLDHTLCRYNLPEASAPIYNSFAQFLVKEKGYDKELLN
VTPEDWDFCCKGLALDLEDGNFLKLANNGTVLRASHGTKMMTPEVLAEAYG
KKEWKHFLSDTGMACRSGKYYFYDNYFDLPGALLCARVVDYLTKLNNQKTT
25 FDFWKDIVAAIQHNYKMSAFKENCGIYFPEIKRDPGRYLHSCPESVKKWLRQL
KNAGKILLITSSHSDYCRLLCEYILGNDFTDLFDIVITNAL.KPGFFSHLPSQRPF
RTLENDEEQEALPSLDKPGWYSQGNAVHLYELLKKMTGKPEPKVVFYFGDSMH
SDIFPARHYSNWETVLIILEELRGDEGTRSQRPEESEPLEKKGKYEGPKAKPLNT
SSKKWGSFFIDSVLGLENTEDSLVYTWSCKRISTYSTIAIPSIEAIAELPLDYKFT
30 RFSSSNSKTAGYYPNPLVLSSDETLISK (SEQ ID NO:233); and/or
TSSHSDYCRLLCEYILGNDFTDLFDIV (SEQ ID NO:234). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. Additionally, another embodiment for this gene is the polynucleotide fragments comprising the following sequence:

35 CCTTAAAAGCTGACATTTATAATTGTGTTGTATAGCAGCAACTATATCCTTC
CAAAAATCAAATGTTTTGACCATTGTCAGTT (SEQ ID NO:230);
CCTTAAAAGCT GACATTATAATTGTGTTGTATAGCA (SEQ ID NO:231);

and/or CTTCCAAAAA TCAAATGTTTTGACCATTGTCAGTT (SEQ ID NO:232). An additional embodiment is the polypeptide fragments encoded by these polynucleotide fragments. This gene maps to chromosome 6, and therefore, may be used as a marker in linkage analysis for chromosome 6.

5 This gene is expressed primarily in prostate and smooth muscle. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostate cancer and cardiovascular disorders. Similarly, polypeptides 10 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate and cardiovascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, and smooth muscle, and 15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of prostate cancer and other disorders. In addition the expression in smooth muscle would suggest a role for this gene product in the treatment and diagnosis of cardiovascular disorders such as hypertension, restenosis, atherosclerosis, stroke, angina, thrombosis, and other aspects of heart disease and respiration.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in endometrial tissue and to a lesser extent in synovium.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endometrial cancer and arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of 35 the above tissues or cells, particularly of the reproductive and skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrial tissue and other reproductive tissue,

and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 130 as residues: Ser-19 to His-24, Pro-36 to Arg-43, Ala-61 to Gly-67, Pro-86 to Ala-95.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of endometrial cancers, as well as reproductive and developmental disorders (fetal deficiencies and other pre-natal conditions). In addition the expression of this gene product in synovium would suggest a role in the detection and treatment of disorders and conditions affecting the skeletal system, in particular the connective tissues (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation).

15

FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene maps to chromosome 6, and therefore, may be used as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in keratinocytes, fetal tissue (especially fetal brain) and leukocytic cell types and tissues (e.g. B-cell, macrophages, Jurkat T-Cell, T cell helper cells, spleen, thymus and lymphoma).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, integument and immune systems, as well as developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., keratinocytes, brain and other tissue of the nervous system, differentiating tissue, leukocytes and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. Expression in keratinocytes would suggest a role for the gene product in the diagnosis treatment of skin disorders such as cancers (melanomas), eczema, psoriasis, wound healing and grafts. In addition the expression in fetal brain might implicate this gene product in the detection and treatment of developmental and neurodegenerative diseases of the brain and nervous system: behavioral or nervous system disorders, such as depression, schizophrenia, Alzheimer's disease, Parkinson's disease, Huntington's disease, mania, dementia, paranoia, addictive behavior and sleep disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 22

Translation product of this gene shares significant homology with the conserved YME1 PROTEIN from *Saccharomyces cerevisiae*, which is a putative ATP-dependent protease thought to regulate the assembly of key respiratory chains within the mitochondria (See Accession No. P32795). Preferred polypeptide fragments comprise the following amino acid sequence:

MKTKNIPEAHQDAFKTGFAEGFLKAQALTQKTNDLRRTRLILFVLLLFGIYGL
LKNPFLSVRFRTTGLDSAIDPVQMKNVTFEHVKGVEEAKQELQEVVEFLKNP
QKFTILGGKLPKGILLVGPPGTGKTLLARA VAGEADVPFYASGSEFDEMFGV
VGASRIRNLFREAKANAPCVIFIDEELDSVGGKRIESPMHPYSRQTINQLLAEMD
GFKPNEGVIIGATNFPEALDNALIRPGRFDMQVTVPRPDVKGRTEILKWYLNK
IKFDXSVDPEIIARGTVGFSGAELENLVNQAALKAAVDGKEMVTMKELGVFQR
25 QNSNGA (SEQ ID NO:235); MKTKNIPEAHQDAFKTGFAEG (SEQ ID NO:236);
PVQMKNVTFEHVKGVEEAKQELQ (SEQ ID NO:237);
SRQTINQLLAEMDGFKPN EGVII (SEQ ID NO:238); and/or
FSGAELENLVNQAALKAAVDGKEM (SEQ ID NO:239). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

30 This gene is expressed primarily in T-cells.
Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an

5 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune 10 disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. Furthermore, the homology of this gene indicates that it may play an important role in disorders affecting metabolism.

15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 23**

This gene is expressed primarily in human chronic synovitis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 20 not limited to, synovial and other inflammatory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the synovial tissue and immune system, expression of this gene at significantly higher or lower levels may be routinely detected 25 in certain tissues and cell types (e.g., synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that the protein product of this gene are useful for study, diagnosis and treatment of inflammatory disorders such as chronic synovitis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 24

This gene is expressed primarily in pituitary, breast cancer, and bone marrow; 35 and to a lesser extent in breast, prostate, uterine cancer and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endocrine, reproductive disorders and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of 5 disorders of the above tissues or cells, particularly of the reproductive, metabolic and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary, mammary tissue, bone marrow, prostate, reproductive tissue, uterus, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, 10 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 134 as residues: Asp-32 to Gln-38, Lys-88 to Ile-97.

15 The tissue distribution indicates that the protein products of this gene are useful for the study, treatment and diagnosis of various endocrine disorders, reproductive diseases and disorders and cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 25

20 The translation product of this gene shares sequence homology with androgen withdrawal apoptosis protein in rat which is thought to be important in programmed cell death. Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

LPMWQVTAFLDHNTVTAQTTWKGLWMSCVVQSTGHMQCKVYDSVLALSTEV
25 QAARALTVSAVLLAFVALFVTLAGAQCTTCVAPGPAKARVALTGGVLYLFCGL
LALVPLCWFAFIVVREFYDPSVPVSQKYELGAXLYIGWAATALLMVGGCLLCC
GAWVCTGRPDLSPVVKYSAPRRPTATGDYDKKNYV (SEQ ID NO:240). This
30 polypeptide is expected to contain multiple transmembrane domains. The extracellular portion of the polypeptide is expected to comprise residues 1-51 of the foregoing amino acid sequence. Therefore, particularly preferred polypeptides encoded by this gene comprise residues 1-51 of the foregoing amino acid sequence. Polynucleotides encoding the foregoing polypeptides are also provided.

This gene is expressed primarily in human adult pulmonary and brain (striatum) tissue and to a lesser extent in thymus, synovium and testis.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, reproductive, metabolic, and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive, 5 nervous, respiratory and metabolic systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, synovial tissue, testis and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, 10 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to androgen withdrawal apoptosis rat gene protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of disorders in which the mechanism 15 controlling programmed cell death is instrumental. This could include reproductive, neurodegenerative, and various metabolic disorders and diseases such as cancer.

FEATURES OF PROTEIN ENCODED BY GENE NO: 26

The translation product of this gene shares homology with both ubiquitin and a 20 G-protein coupled receptor TM3 consensus polypeptide (see Genbank accession Nos. gnl|PIDle331456 (AJ000657) and R50664, respectively). Preferred polypeptides encoded by this gene comprising the following amino acid sequence: LHYFALSFVILTEICLVSSGMGF (SEQ ID NO:241); QLRNGIPPGRKALFCGKPR LFTLGQQRTCA (SEQ ID NO:242); and/or 25 WSGLWVTTWNGSSGERTPSPWRRK RASQSAGRIASWMSF (SEQ ID NO:243). An additional embodiment is polynucleotides encoding these polypeptides. This gene maps to chromosome 1, and therefore, may be used as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in activated T cells and to a lesser extent in 30 CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hemopoietic disorders. Similarly, polypeptides and 35 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hemopoietic and immune system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other blood cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell

5 sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 136 as residues: Thr-15 to His-21, Gly-30 to Lys-39, Arg-113 to Met-118, Arg-178 to Ala-187.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex vivo culture, bone marrow

15 transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. Furthermore, the homology to G-coupled proteins as well as to ubiquitin may implicate this gene as being important in regulation of gene expression and protein

20 sorting - both of which are vital to development and would healing models. Therefore, the gene may provide utility in the diagnosis, prevention, and/or treatment of various developmental disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 27

25 This gene is expressed primarily in activated T cells and to a lesser extent in fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

30 not limited to, immune, developmental and metabolic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and metabolic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

35

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study and treatment of diseases and disorders of the immune, metabolic, and endocrine systems; such as renal diseases and T cell dysfunctions. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 28

The translation product of this gene shares sequence homology with Cystatin-related epididymal specific protein in mouse which is thought to be important in

reproductive system function/regulation (See Genbank accession no.bbs118813). Based on the structural similarity between these proteins, the translation product of this clone, hereinafter "Cystatin G", is expected to share biological activities with cystatin related proteins and other cysteine protease inhibitors. Such activities are known in the art and are described elsewhere herein. Preferred polypeptides encoded by this gene comprising the following amino acid sequence:

MPRCRWLSLILLTIPLALVARKDPKKNETGVLRKLKPVNASNANVKQCLWFA
MQEYNKESEDKYVFLVVKTLQAQLQVTNLLEYLIDVEIARSDCRKPLSTNEICAI

QENSKLKRKLSCSFLVGALPWNGEFTVMEKKCEDA (SEQ ID NO:246);
ARKDPKKNETGVLRKLKPVNASNANVKQCLWFAMQEYNKESEDKYVFLVVK

25 TLQAQLQVTNLLEYLIDVEIARSDCRKPLSTNEICAIQENSKLKRKLSCSFLVGA
LPWNGEFTVMEKKCEDA (SEQ ID NO:248);

CLWFAMQEYNKESEDKYVFLVVKTLQAQLQVTNLLEYLIDVEIARSDCRKPLST
NEICAIQENSKLKRKLSCSFLVGALPWNGEFTVMEKKC (SEQ ID NO:247);

EYNKESEDKYVFLV (SEQ ID NO:244); and/or IDVEIARSDCRKPL (SEQ ID

30 NO:245). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. Preferred cystatin polypeptide fragments are shown to be active in the following assays: The methods used for active site titration of papain, titration of the molar enzyme inhibitory concentration in cystatin G preparations, and for determination of equilibrium constants for dissociation (Ki) of complexes between

35 cystatin G and cysteine peptidases are described in detail in Hall et al., Biochem. J., 291:123-29 (1993) and Abrahamson, Methods Enzymol., 244:685-700 (1994), both of which are hereby incorporated herein by reference. The enzymes used for equilibrium

assays are papain (EC 3.4.22.2; from Sigma, St Louis, MO) and cathepsin B (EC 3.4.22.1; from Calbiochem, La Jolla, CA). The fluorogenic substrate used was Z-Phe-Arg-NHMec (10 mM; from Bachem Feinchemikalien, Bubendorf, Switzerland) and the assay buffer was 100 mM Na-phosphate buffer (pH 6.5 and 6.0 for papain and 5 cathepsin B, respectively), containing 1 mM dithiothreitol and 2 mM EDTA. Steady state velocities are measured and K_i values were calculated according to Henderson, Biochem J., 127:321-333 (1972), incorporated herein by reference. Corrections for substrate competition are made using K_m values of 150 μ M for cathepsins B (Barrett and Kirschke, Methods Enzymol., 80:535-561 (1981) and 60 μ M for papain (Hall et 10 al., Biochem. J., 291:123-29 (1992)), both of which are hereby incorporated herein by reference.

This gene is expressed primarily in human testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 15 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this 20 gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testis and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily 25 fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 138 as residues: Arg-21 to Thr-29.

The tissue distribution and homology to cystatin-related epididymal specific protein-mouse indicates that polynucleotides and polypeptides corresponding to this 30 gene are useful for study, diagnosis and treatment of reproductive diseases and disorders. Cysteine proteinase inhibitors of the cystatin superfamily are ubiquitous in the body and are generally tight-binding inhibitors of papain-like cysteine proteinases, such as cathepsins B, H, L, S, and K (for review, see Ref. 1). They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled proteolysis and tissue damage. Cysteine 35 proteinase activity can normally not be measured in body fluids, but can be detected extracellularly in conditions like endotoxin-induced sepsis (2), metastasizing cancer (3), and at local inflammatory processes in rheumatoid arthritis (4), purulent bronchiectasis

(5) and periodontitis (6), which indicates that a tight cystatin regulation is a necessity in the normal state. A deficiency state in which the levels of the intracellular cystatin, cystatin B, are lowered due to mutations has recently been shown to segregate with a form of progressive myoclonus epilepsy (7), which points to additional specialized

5 functions of cystatins. Moreover, results showing that chicken cystatin inhibits polio virus replication (8), human cystatin C inhibits corona- and herpes simplex virus replication (9,10), and human cystatin A inhibits rhabdovirus-induced apoptosis (11) in cell cultures indicates that cystatins play additional roles in the human defense system. The cystatins constitute a superfamily of evolutionary related proteins, all composed of

10 at least one 100-120 residue domain with conserved sequence motifs (12). The previously well characterized single-domain human members of superfamily could be grouped in two protein families. The Family 1 members, cystatins (or stefins) A and B, contain approximately 100 amino acid residues, lack disulfide bridges, and are not synthesized as preproteins with signal peptides. The Family 2 cystatins (cystatins C, D,

15 S, SN, and SA) are secreted proteins of approx. 120 amino acid residues (Mr 13,000-14,000) and have two characteristic intrachain disulfide bonds. Recently, we identified an additional human cystatin superfamily member by EST1 sequencing in epithelial cell derived cDNA libraries which we named cystatin E (13). The same cystatin was

20 independently discovered by differential display experiments as a mRNA species down-regulated in breast tumor tissue, but present in the surrounding epithelium and reported under the name cystatin M (14). Cystatin E/M is an atypical, secreted low-Mr cystatin in that it is a glycoprotein and just shows 30-35% sequence identity in alignments with the human Family 2 cystatins, which shows that additional cystatin families are yet to be identified (13). The cystatin E/M gene has been localized to chromosome 2 (15),

25 whereas all human Family 2 cystatin genes are clustered on the short arm of chromosome 20 (16), which further stresses that cystatin E/M is just distantly related to the other secreted human low-Mr cystatins.

FEATURES OF PROTEIN ENCODED BY GENE NO: 29

30 The translation product of this gene shares sequence homology with the leukocyte-associated Ig-like receptor-1, putative inhibitory receptor which is thought to be important in regulation of various physiological functions (See Accession No. gil2352941 (AF013249). Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

35 DSPDTEPGSSAGPTQRPSDNHNEHAPASQGLKAEHLYILIGVS (SEQ ID NO:249); HRQNQIKQGPPRSKDEEQKPQQRPDLAVDVLERTADKATVNGL PEKDRETDTSALAAGSSQEVTYAQLDHWALTQRTARAVSPQSTKPMAESITYAA

VARH (SEQ ID NO:250);

MSPHPTALLGLVLCLAQTIHTQEEDLPRPSIAEPTVIPLGSHVTFCRGPVGV
QTFRLERESRSTYNDTEDVSQASPSEARFRIDSSEGNAQPYRCIYYKPPKW
SEQSDY (SEQ ID NO:251); TALLGLVLCLAQTIHTQE (SEQ ID NO:252);

5 LPRPSIAEPTVI (SEQ ID NO:253); CRGPVGVQTFRLE (SEQ ID NO:254);
and/or VLERTADKATVNGLPEKDRETDTSALAAGSS (SEQ ID NO:255).
Additional embodiments of the invention include polynucleotides encoding these
polypeptides.

This gene is expressed primarily in macrophages and T-cells and to a lesser
10 extent in human fetal heart.

Therefore, polynucleotides and polypeptides of the invention are useful as
reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis of diseases and conditions which include, but are
not limited to, developmental, inflammatory, and immune disorders. Similarly,
15 polypeptides and antibodies directed to these polypeptides are useful in providing
immunological probes for differential identification of the tissue(s) or cell type(s). For a
number of disorders of the above tissues or cells, particularly of the growth and
inflammatory systems, expression of this gene at significantly higher or lower levels
may be routinely detected in certain tissues and cell types (e.g., macrophages, T-cells
20 and other cells and tissue of the immune system, heart, and cancerous and wounded
tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or
another tissue or cell sample taken from an individual having such a disorder, relative to
the standard gene expression level, i.e., the expression level in healthy tissue or bodily
fluid from an individual not having the disorder. Preferred epitopes include those
25 comprising a sequence shown in SEQ ID NO. 139 as residues: His-20 to Arg-28, Glu-
61 to Val-74, Ser-78 to Ala-84, Lys-105 to Ser-117.

The tissue distribution and homology to putative inhibitory receptor indicates
that polynucleotides and polypeptides corresponding to this gene are useful for the
study, diagnosis and treatment of functional disorders of the developing fetal heart;
30 including circulatory and vascular; and inflammatory disorders. In addition expression
in macrophages and lymphocytes indicates a role in the treatment/detection of immune
disorders including disorders such as arthritis, asthma, immune deficiency diseases
such as AIDS, and leukemia.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 30

The translation product of this gene shares sequence homology with erythroid
cell specific transcription factor- murine which is thought to be important in normal

physiological function of erythroid cells. In addition, the translation product of this gene also shares homology with the conserved 3-phosphoglycerate dehydrogenase gene which is essential component of metabolic biosynthetic pathways. Preferred polypeptides comprise the following amino acid sequence:

5 MNTPNGNSLSAAELTCGMIMCLARQIPQATASMKDGKWERKKFMGTELNGK
TLGILGLGRIGREVATRMQSGMKTIGYDPIISPEVSASFGVQQLPLEEIWPLCDF
ITVHTPLLPSTTGLLNDNTFAQCKKGVRVNCARGGIVDEGALLRALQSGQCA
GAALDVFTEEPPRDRALVDHENVISCPHLGASTKEAQSRCGEEIAVQFVDMVK
GKSLTGVVNAQALTSAFSPHTKPWIGLAEALGTLMRAWAGSPKGTIQVITQGT

10 SLKNAGNCNSPAVIVGLLKEASKQADVNLVNAKLLVKEAGLNVTTSHPAAAPG
EQGFGECLLAVALAGAPYQAVGLVQGTTPVLQGLNGAVFRPEVPLRRDLPILL
FRTQTSDPAMLPTMIGLLAEAGVRLLSYQTSLVSDGETWHVMGISSLLPSLEAW
KQHVTEAFQFHF (SEQ ID NO:256); MAFANLRKVLISDSLDPCCRKILQ (SEQ ID
NO:257); GGLQVVEKQNL SKEELIA (SEQ ID NO:258);

15 MCLARQIPQATASMKDGKWERKKFMGTEL (SEQ ID NO:259);
ALTSAFSPHTKPWIGLAEALGTLMRAWAG (SEQ ID NO:260); and/or
EVPLRRDLPILLFRTQTSDPAMLPTMIGLLAEAGVR (SEQ ID NO:261). Also
preferred are polynucleotide fragments encoding these polypeptides. This gene maps to
chromosome 1, and therefore, may be used as a marker in linkage analysis for
20 chromosome 1.

This gene is expressed primarily in IL-1 induced smooth muscle and fetal kidney and to a lesser extent in myeloid progenitor cell line and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a
25 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune, hemopoietic, and cardiovascular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hemopoietic and
30 immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, kidney, myeloid progenitor cells, bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene
35 expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 140 as residues: Met-1 to Asn-7, Met-33 to Lys-42,

Asn-123 to Cys-130, Glu-169 to Asp-174, Ser-192 to Gly-201, Thr-266 to Asn-273, Pro-318 to Phe-323.

The tissue distribution and homology to erythroid cell specific murine transcription factor indicates that polynucleotides and polypeptides corresponding to 5 this gene are useful for study, diagnosis and treatment of disorders and diseases involving the hemopoietic and immune systems; the maturation of progenitor cells; and the development of various smooth muscle tissues (heart, etc.). In addition, homology to a key biosynthetic protein implicates this the protein product of this gene as being important in metabolism. Therefore, the protein may show utility in the diagnosis, 10 prevention, and/or treatment of metabolic disorders and conditions.

FEATURES OF PROTEIN ENCODED BY GENE NO: 31

This gene is expressed primarily in human adult testes.

Therefore, polynucleotides and polypeptides of the invention are useful as 15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive disorders, particularly of the male genitalia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a 20 number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the 25 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 141 as residues: Met-1 to Pro-8, Ser-45 to Thr-50.

The tissue distribution indicates that polynucleotides and polypeptides 30 corresponding to this gene are useful for the study, diagnosis, treatment, and possibly prevention of various male reproductive disorders and diseases including male impotence, failed libido and male secondary sex characteristics, infertility, and testicular cancer.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in human adult testis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive disorders and cancers of the male reproductive system.

5 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testis and other reproductive
10 tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study, diagnosis, treatment, and possibly prevention of various male reproductive disorders and diseases including male impotence, failed libido and male secondary sex characteristics, infertility, and testicular cancer.

20

FEATURES OF PROTEIN ENCODED BY GENE NO: 33

The translation product of this gene shares homology to the W09D10.1 protein of *Caenorhabditis elegans*. In addition, the gene also shares homology with the human protein hRIP, a protein known to be critical for HIV replication (See Accession

25 Nos.gnl|PID|e1186472 and W12713). Preferred polypeptides encoded by this gene comprise the following amino acid sequence:

MDLLGLDAPVACSIANSKTSNTLEKDLLLASVPSPSSSGSRKVVGSMPTAGSA
GSVPENLNLFPEPGSKSEEIGKKQLSKDSILSLYGSQTXQMPTQAMFMAPAQMQ
AYPTAYPSFPGVTPPNSIMGSMMPPPVGVMVAQPGASGMVAPMAMPAGYMG
30 MQASMMGVPNGMMTTQQAGY MAGMAAMPQT VYGVQPAQQLQWNLTQMTQ
QMAGMNFY GANGMMNYGQSMSGGNGQAANQTLSPQMWKFGTRFLANLL
EDNKFCADCQSKGPRWASWNIGVFICIRCA XIHRNLGVHISRVKS VNL DQWTQ
VQIQC (SEQ ID NO:267); MQXMGNGKANRLYEAYLPETFRRPQIDPAVEGFIR
DXYE (SEQ ID NO:268); EEDNKFCADCQSKGPRWASWN (SEQ ID NO:263);
35 GVFICIRCA XIHR NLGVHIS (SEQ ID NO:264); and/or SVNLDQWTQVQIQC MQX
MGNGKA (SEQ ID NO:265). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in lymphoid tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and inflammatory disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, hematopoietic and inflammatory, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue and other tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 143 as residues: Cys-21 to Trp-28.

The tissue distribution indicates that the protein products of this gene are useful for study, diagnosis and treatment of various immune disorders and diseases, including self-recognition and rejection functions of the immune system, hematopoietic disorders, and inflammatory disorders. Homology to the W09D10.1 of *C.elegans* and the hRIP implicates this gene as playing a role as an essential receptor for host-viral interactions including, but not limited to retroviral infections such as AIDS.

FEATURES OF PROTEIN ENCODED BY GENE NO: 34

25 The translation product of this gene shares homology to an *Arabidopsis thaliana* recombination and DNA-damage resistance/repair protein (See Accession No.gil166694). Preferred polypeptides encoded by this gene comprise the following amino acid sequence:
KYGKVGKCVIFEIPGAPDDEAVRIFLEFERVERVESAIKA VVDLNGRYFGGRVVKAC
30 FYNLDKFRVLDLA (SEQ ID NO:269); KAVDLGRYFGGR (SEQ ID NO:270); and/or EAVRIFFRE (SEQ ID NO:271). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in ovarian and other cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly of the female reproductive system. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely 5 detected in certain tissues and cell types (e.g., ovaries and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the 10 disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 144 as residues: Thr-11 to Trp-19, Ala-40 to Gln-47, Lys-58 to Arg-66, Asp-98 to Lys-110, Arg-114 to Glu-121.

The tissue distribution in tumors of ovarian origins combined with the 15 homology to a known DNA damage repair enzyme indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of tumors. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

FEATURES OF PROTEIN ENCODED BY GENE NO: 35

20 Translation product of this gene shares homology with human stomatin, intestinal surface antigens, as well as protein F30A10.5 of *Caenorhabditis elegans* (See Accession No.gnllPIDle276130). Preferred polypeptides encoded by this contig comprise the following amino acid sequence: RMGRFHRILEPGLNLIPVLDRIYVQ SLKEIVINVPEQSAVTLDNVTLQIDGVLYLRIMDPYKASYGVEDPEYAVTQLAQQT 25 TMRSELGKLSLDKVFRERESLNASIVDAINQAADCWGIRCLR YEIKDIHVPPRV KESMQMQVEAERRKRATVLESEGTRRESAINVAEGKKQAQILASEAEKAEQINQA AGEASAVLAKAKAKAEAIRILAAALTQHNGDAAASLTVAEQYVSAFSKLAKDS NTILLPSNPGDVTSMSVAQAMGVYVGALTAKAPVPGTPDSLSSGSSRDVQGTDASL DEELDRVKMS (SEQ ID NO:272); ASYGVEDPEYAVTQLAQTT MRSELGK (SEQ 30 ID NO:273); MQMQVEAERRKRATVLESEGTRRESAIN (SEQ ID NO:274); LTVAEQYVSAFSKLAKDSNTILLPSN (SEQ ID NO:275), and/or LLGATAPLVSLVPEVAAA VGNAGARGAXHWGPFAEGLSTGFWPRSARASSGL PRNTVVLFVPQQEAWVVE (SEQ ID NO:276). Polynucleotides encoding these polypeptides are also provided.

35 This gene is expressed primarily in activated T-cells and to a lesser extent in other cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to 5 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and cancerous and 10 wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 145 as residues: Arg-23 to Pro-33, 15 Pro-184 to Ser-189, Ala-196 to Arg-201, Glu-208 to Ser-213, Glu-230 to Ile-237, Gly-326 to Leu-331, Gly-334 to Gln-340.

The tissue distribution indicates that the protein products of this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, 20 pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, 25 inflammation, allergy, immunodeficiency etc. In addition, the homology to known intestinal antigens may suggest that the protein is important in the diagnosis, treatment, and/or prevention of gastrointestinal disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 36

Translation product of this gene has homology to a human estrogen receptor 30 variant from human breast cancer. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: RMWRNGTHFWECKIVQLWK TVWWFPRKLSIELPENLAILIGTYFK (SEQ ID NO:277); and/or LKRHFPKEANK HVKRCSTSLDIREIQIKIKMRY (SEQ ID NO:278). Polynucleotides encoding these polypeptides are also provided.

35 This gene is expressed primarily in ulcerative colitis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, intestinal ulcers, inflammatory conditions and cancers, particular of the breast. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the gastrointestinal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon and other gastrointestinal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in colon and breast origins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of tumors or other conditions within these tissues, in addition to other tumors where expression has been indicated. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues.

20 FEATURES OF PROTEIN ENCODED BY GENE NO: 37

This gene is expressed primarily in epithelial cells. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers and skin disorders, particularly melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin and other epithelia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 147 as residues: Met-1 to Tyr-6.

The tissue distribution in epithelial tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of

tumors of this tissue. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 38

This gene is expressed primarily in adult retina.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases of the eye. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the eye, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 148 as residues: Cys-14 to Lys-21.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of the eye.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 39

This gene is expressed primarily in bone marrow and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hemopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hemopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard

gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of the 5 hemopoietic system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 40

This gene is expressed primarily in lymph node, fetal liver and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hemopoietic diseases and disorders of the CNS. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of 15 disorders of the above tissues or cells, particularly of the hemopoietic and CNS, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue and other tissue of the immune system, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal 20 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful 25 for the diagnosis and treatment of cancer and other proliferative disorders. Expression in embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation or cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the 30 maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. In addition, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, 35 obsessive compulsive disorder, panic disorder, and autism. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental

disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 41

5 The translation product of this gene shares sequence homology with fibropellin and epidermal growth factors which are thought to be important in growth and regeneration of epidermal cells (See Genbank Accession Nos. W11719 and gil310660). Preferred polypeptides comprise the following amino acid sequence:

GTRPGESHANDLECSGKGKCTKPSEATFSCTCEEQYVGTFC EYDACQRKPC

10 QNNASCIDANEKQDGSNFTCVCLPGYTGECLQSKIDYCILDPCRNGATCISSLS GFTCQCPEGYFGSACEEKVDP CASSPCQNNGTCYVDGVHFTCNCSPGFTGPTC AQLIDFCALSPCAHGTCSR VGT SYKCLCDPGYHGLYCEEYNECLSAPCLNAA TCRDLVNGYECVCLAEYKGTHCELYKDPCANVSLNGATCDS DGLNGTCICA PGFTGEECDIDINECD SNPCHHGGSCLDQPNGYNCHCPHG WVGANCEIHLQW

15 KSGHMAESLTN (SEQ ID NO:279); GKCTKPSEATFSCTCEEQYVGTFC (SEQ ID NO:280); CAHG TCRSVGT SYKCLCDPGYH (SEQ ID NO:281); and/or CANVSLNGATCDS DGLNG TCICAPGFTGEECD (SEQ ID NO:282).

Polynucleotides encoding these polypeptides are also provided.

20 This gene is expressed primarily in brain and kidney and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the neural and renal systems, particularly growth disorders 25 such as cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural and renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other 30 tissue of the nervous system, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution and homology to epidermal growth factor indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of growth disorders especially in the neural and renal systems. In

addition, polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, and autism.

5 In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system

FEATURES OF PROTEIN ENCODED BY GENE NO: 42

10 This gene is expressed primarily in brain, kidney and stromal cells. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the CNS and hemopoietic system. Similarly, polypeptides 15 and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hemopoietic, renal and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the 20 nervous system, kidney, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those 25 comprising a sequence shown in SEQ ID NO. 152 as residues: Lys-71 to Trp-76, Glu-99 to Gly-108, Arg-142 to Ser-149.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's 30 Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, and autism. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. In addition, polynucleotides and polypeptides corresponding to 35 this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include

bone marrow cell ex vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product is thought to be involved in lymphopoiesis, therefore, it can be used in immune disorders to modulate infection, inflammation, allergy, immunodeficiency, etc.

5

FEATURES OF PROTEIN ENCODED BY GENE NO: 43

The preferred polypeptide encoded by this gene comprise the following amino acid sequence: MAQNLKDLAGRLPAGPRGMGTALKLLLGAGAVAYGVRESVFT VEGGHRAIFFNRIGGVQQDTILAEGLHFRIWFQYPIIYDIRARPRKISSPTGSKD 10 LQMVNISLRVLSRPNAQELPSMYQRLGLDYEERVLPSIVNEVLKSVVAKFNASQ LITQRAQVSLLIRRELTERAKDFSLILDDVAITELSFREYTAAVEAKQVAQQEAQ RAQFLVEKAKQEQRQKIVQAEGEAEAAMLGEALSKNPGYIKLRKIRAAQNS 15 KTIATSQNRIYLTADNLVNLQDESFRGSDSLIKGKK (SEQ ID NO:283). The gene product above share sequence similarity with prohibitin. Thus, these polypeptides are expected to share biological activities with prohibitin. Such activities are known in the art and discussed elsewhere herein.

This gene is expressed primarily in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or 25 lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an 30 individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 153 as residues: Ala-85 to Ser-91, Pro-93 to Asp-98, Glu-167 to Lys-173, Gln-205 to Ala-210.

The tissue distribution and structural similarity to prohibitin indicates that the protein products of this gene are useful for the detection/treatment of neurodegenerative 35 disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, and autism. In addition, the gene or gene product

may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, and/or disorders of the cardiovascular system.

5 FEATURES OF PROTEIN ENCODED BY GENE NO: 44

The translation product of this gene shares sequence homology with the F44G4.1 gene of the *c. elegans* genome which has no known function (See Accession No.gnllPIDle236516). The translation product of this gene also shares sequence homology with the human torsionA and torsionB gene products, a gene candidate for 10 the Torsion Dystonia disease locus (See Accession Nos gil2358279 (AF007871) and gil2358281 (AF007872)). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: KALALSFHGWGSGTGKNFV (SEQ ID NO:284); NLIDYFIPFLPLEYRHVRRLCAR (SEQ ID NO:285); NLIDYFIPFLPL 15 EYRHVRRLC (SEQ ID NO:286); CHQTLFIFDEAEKLHPGLLEVLGPHL (SEQ ID NO:287); and/or PEKALALSFHGWGSGTGKNFVA (SEQ ID NO:288). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in tonsils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, such as tonsilitis or adenoiditis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at 25 significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., tonsils, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the 30 disorder.

The tissue distribution and homology to F44G4.1 gene of the *c. elegans* genome indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and detection of conditions affecting the tonsils. The tonsils have not been thoroughly studied and the actual function of this organ is not known, 35 but this gene could be used in determining what may trigger tonsilitis. Especially in children, where the tonsils seem to be most active. Furthermore, due to the homology

of this gene, it may display potential utility in the detection, diagnosis, and/or treatment for Torsion Dystonia disease.

FEATURES OF PROTEIN ENCODED BY GENE NO: 45

5 Has exact sequence homology on the nucleotide level as Human HepG2 3' region cDNA, but the function of this gene is not known.

This gene is expressed primarily in osteoclastoma stromal cells and to a lesser extent in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, leukemia and bone disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the haemolymphoid system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of diseases such as leukemia.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed primarily in activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders, including leukemia and allergies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lymphoid system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hemopoietic cells, bone marrow, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 156 as residues:

5 Met-1 to Gly-7.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment in tissue repair and modeling since monocytes engage the synthesis and secretion of many cytokines which are soluble proteins that regulate highly diverse aspects of cellular biology. Monocytes are 10 also important in the fact that their expression of Major Histocompatibility Factor II (MHCII) enable them to select and stimulate the appropriate lymphocytes to combat specific antigens in the blood. Since the gene is expressed in cells of lymphoid origin, the natural gene product may be involved in immune functions. Therefore it may be also used as an agent for immunological disorders including arthritis, asthma, immune 15 deficiency diseases such as AIDS, and leukemia.

FEATURES OF PROTEIN ENCODED BY GENE NO: 47

Translation product of this gene has homology to the Na⁺/H⁺-exchanging protein: Na⁺/H⁺ antiporter in *Methanobacterium thermoautotrophicum* as well as the 20 Na⁺/H⁺ antiporter cdu2' in *Clostridium difficile* (See Accession Nos. gil2621849 (AE000854) and pirJC5343|JC5343, respectively). Thus, it is likely that this gene has similar Na⁺/H⁺ antiporter activity. One embodiment for this gene are polypeptide fragments comprising the following amino acid sequence:

NLKEKIFISFAWLPKATVQAAIG (SEQ ID NO:289) and/or

25 WLPKATVQAAIGSVALD (SEQ ID NO:290). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in osteoclastoma cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 30 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, osteoporosis, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lymphoid and skeletal systems, expression of this 35 gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell

sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 157 as residues: His-35 to Gln-43.

5 The tissue distribution predominantly in osteoclastoma cells (the site of hematopoiesis) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of bone related diseases including osteoporosis, osteopetrosis and leukemia. Furthermore, its homology to known transporter proteins may suggest the protein is useful in the diagnosis, treatment, and 10 prevention of various developmental and metabolic disorders, particularly those based upon ion and proton transport.

FEATURES OF PROTEIN ENCODED BY GENE NO: 48

This gene is expressed primarily in amygdala and to a lesser extent in amniotic 15 cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, depression and other emotional behavioral problems. Similarly, 20 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and tissues of the nervous system, and 25 tissues of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid or amniotic fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of mental problems associated with emotional behavior and neurodegenerative states such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorders, and 35 depression. The amygdala processes sensory information and relays this to other areas of the brain including the endocrine and autonomic domains of the hypothalamus and the brain stem. In addition, expression of this protein in amniotic cells suggests that

this protein would be useful in the diagnosis, prevention, and/or treatment of various developmental and/or reproductive system disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 49

5 This gene is expressed primarily in stromal cells. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, leukemia and other cancers and disorders deriving from hematopoietic

10 cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the lymphoid system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., haematopoietic tissues, and

15 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or lymph fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex vivo culture, bone marrow

25 transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc.

30 FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene maps to chromosome 9, and therefore, may be used as a marker in linkage analysis for chromosome 9.

This gene is expressed primarily in tumors, particularly skin and adrenal gland tumors, and to a lesser extent in bone marrow stromal cells and activated T cells.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, cancer; hematopoietic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, adrenal gland, and 5 immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endocrine glands, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 10 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 160 as residues: Glu-13 to Arg-22, Ser-58 to Trp-63.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of cancer. Elevated 15 levels of expression of this gene in a variety of tumors suggest that it may play a role in cell proliferation, the induction of angiogenesis, destruction of the basal lamina, or a variety of other physiological processes that support the growth and development of tumors and cancer. Alternatively, its expression in the hematopoietic compartment, particularly in the bone marrow stroma and by activated T cells suggest that it may 20 represent a soluble factor capable of influencing a variety of hematopoietic lineages. Therefore, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of blood cells.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 51

This gene is expressed primarily in benign human breast tissue. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 30 not limited to, breast cancer and other female reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast and reproductive tissues, expression of this gene at significantly higher or lower levels may 35 be routinely detected in certain tissues and cell types (e.g., breast tissue, secretory/ductile organs, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid or milk) or another tissue or cell

sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or diagnosis of breast cancer. Alternately, this protein may play an important role in lactation or represent a critical component secreted into the milk, which may have an important function in the immunoprotection, health, and/or nourishment of the infant upon breastfeeding. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues

FEATURES OF PROTEIN ENCODED BY GENE NO: 52

Translation product of this gene has homology with the conserved human ring finger proteins (See Accession No.gnll|PIDle351238 (AJ001019)) which are thought to be important in facilitating and regulating signal transduction pathways in eukaryotic cells. One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: HDRTMQDIVYKLVPGLQE (SEQ ID NO:291) and/or FASHDRTM QDIVYKLVPGLQE (SEQ ID NO:292). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in adult whole brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disorders; Schizophrenia; Alzheimer's; tumors of a brain or neuronal cell origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the CNS and/or peripheral nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 162 as residues: Phe-39 to Gly-44.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative

disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, considering the homology to the conserved ring finger proteins may suggest that the gene or gene product may also play 5 a role in the treatment and/or detection of developmental disorders associated with the developing embryo.

FEATURES OF PROTEIN ENCODED BY GENE NO: 53

Translation product of this gene shares homology with the human conserved 10 Lst-1 gene product, a member of the TNF family of proteins (See Accession No.gil1127546). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: LVLSLGAWGWPSTCLWW (SEQ ID NO:293). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

15 This gene is expressed primarily in human 6-week old embryo. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, abnormal cell proliferation; defects in terminal tissue differentiation.

20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., proliferating and differentiating tissues,

25 and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid or amniotic fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or diagnosis of fetal disorders. Alternately, expression within embryonic tissues may reflect a role for this protein in proliferating cells. In such an event, this gene product may be useful in the treatment or diagnosis of abnormal cell proliferation, such as that involved in cancer.

35 Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis involved in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation, and could again be useful in cancer therapy.

FEATURES OF PROTEIN ENCODED BY GENE NO: 54

This gene is expressed primarily in human epithelioid sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as

5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, epithelial sarcoma; tumors of an epithelial cell origin including the underlying integument. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification 10 of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin and epithelial tissue layers, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., epithelial cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or 15 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 164 as residues: Met-1 to Tyr-6, Thr-24 to Cys-36.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or diagnosis of epithelial cancer. This gene product displays enhanced expression in epithelial cell sarcoma, and thus may be involved in cell proliferation, apoptosis, or in the control of angiogenesis.

FEATURES OF PROTEIN ENCODED BY GENE NO: 55

25 This gene is expressed primarily in endometrial tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endometrial cancer including other cancers of the female reproductive 30 system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endometrium and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., 35 endometrial tissue as well as other tissues of the female reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers,

5 particularly those of the endometrium and other reproductive organs. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tumors and tissues

FEATURES OF PROTEIN ENCODED BY GENE NO: 56

10 This gene is expressed primarily in metastatic melanoma and to a lesser extent in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 15 not limited to, cancer of the integument system, particularly melanoma, as well as within the developing pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at 20 significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cells capable of forming melanin, epithelia, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or pulmonary surfactant) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the 25 expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 166 as residues: Asp-20 to Lys-25.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer, particularly 30 melanoma and more particularly, metastasizing melanomas. In addition, the tissue distribution also indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression in embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation or cellular 35 division.

FEATURES OF PROTEIN ENCODED BY GENE NO: 57

This gene is expressed primarily in T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, lymphomas and other immune derived cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 167 as residues: Met-1 to Asn-7.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of lymphomas, particularly T cell lymphomas, and other cancers. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. Thus, this gene may be useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells.

FEATURES OF PROTEIN ENCODED BY GENE NO: 58

This gene maps to chromosome 7, and therefore is useful in linkage analysis as a marker for chromosome 7.

This gene is expressed primarily in brain and to a lesser extent in spinal cord.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, CNS and PNS diseases and disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain, spinal cord and other tissue of the nervous system, and

5 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 168 as residues:

10 Tyr-14 to Ala-30.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, panic disorder, and autism.

FEATURES OF PROTEIN ENCODED BY GENE NO: 59

Translation product of this gene shares homology to the conserved *C. elegans* protein FER-1 (See Accession No.gil1373333). One embodiment for this gene is the

20 polypeptide fragments comprising the following amino acid sequence:

QGKLQMWMVVDVFPKSL (SEQ ID NO:294); PPFNITPRKAKKYYLR (SEQ ID NO:295); KTDVHYRSLDGEGNFNWRF (SEQ ID NO:296); and/or PRLIIQIWDNDKFSLDDY LGFLELDL (SEQ ID NO:297). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments.

25 This gene is expressed primarily in synovial fibroblasts and to a lesser extent in synovial hypoxia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

30 not limited to, synovial inflammation and other diseases of the joints. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the synovium, expression of this gene at significantly higher or lower levels may be routinely detected

35 in certain tissues and cell types (e.g., synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases affecting the synovium of the joints, such as rheumatoid arthritis, osteoarthritis, other inflammatory conditions affecting the joints, as well as in the detection and treatment of disorders and conditions affecting the skeletal system, in particular the connective tissues (e.g. trauma, tendonitis, chondromalacia and inflammation). Furthermore, the homology to a conserved *C.elegans* protein may suggest protein is important in human development and thus is beneficial in the diagnosis, prevention, and treatment of developmental disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 60

This gene is expressed primarily in endothelial cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammation and other disorders of the integument, in addition to neurodegenerative and nervous system disorder, such as stroke. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endothelial, circulatory, and nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 170 as residues: Ser-4 to Gly-13.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory diseases primarily mediated through endothelial cells, such as sepsis, inflammatory bowel disease, psoriasis, and Crohn's disease, as well as for stroke. Alternatively, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and

behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing 5 embryo, or disorders of the cardiovascular system.

FEATURES OF PROTEIN ENCODED BY GENE NO: 61

This gene is expressed primarily in fetal brain.

Therefore, polynucleotides and polypeptides of the invention are useful as 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, CNS and PNS disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of 15 the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., developing and differentiating tissues, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or amniotic fluid) or another tissue or cell sample 20 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neural disorders 25 such as Alzheimer's disease, depression, paranoia, schizophrenia, autism, and particularly developmental brain disorders..

FEATURES OF PROTEIN ENCODED BY GENE NO: 62

Translation product of this gene shares homology with a conserved 4-30 nitrophenylphosphatase from *Schizosaccharomyces pombe* (See Accession No. gil1938421). One embodiment for this gene is the polypeptide fragments comprising the following amino acid sequence: AVMIGDDCRDDVGGA (SEQ ID NO:298), and/or 35 ILVKTGKYRASDEEKIN (SEQ ID NO:299). An additional embodiment is the polynucleotide fragments encoding these polypeptide fragments. This gene maps to chromosome 18, and therefore, may be used as a marker in linkage analysis for chromosome 18.

This gene is expressed primarily in endometrial tumors and to a lesser extent in leukemia and lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly of the immune and hematopoietic systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endometrium and white blood cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrial and/or proliferating tissues, and cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 172 as residues: Val-19 to Cys-24.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection, diagnosis, and treatment of cancers, particularly those cancers affecting endometrial tissues and the lymphatic system. In addition, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. The uses include bone marrow cell ex vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. Furthermore, homology to a conserved *S.pombe* protein may suggest protein is important in development. Therefore, protein may be beneficial in the diagnosis, prevention, and treatment of developmental disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 63

The translation product of this gene shares sequence homology with ribosomal releasing factor which is thought to be important in protein synthesis.

This gene is expressed primarily in pancreatic tumors, placenta, testis, ovarian cancer, adipocytes, spleen, and fetal liver and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for diagnosis of a number of diseases and conditions such as immune-
5 diseases, cardiovascular and endocrine diseases and others. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, cardiovascular system, digestive system and reproductive system. expression of this
10 gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pancreas, testis and ovary and other reproductive tissue, adipocytes, spleen, liver, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the
15 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 173 as residues: Glu-36 to His-41, Thr-57 to Thr-70, Glu-87 to Met-92, Lys-100 to Lys-105, Ala-197 to Ser-227.

The tissue distribution and homology to ribosomal releasing factor indicates that
20 polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of many diseases, especially cancers and immuno-related diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 64

The translation product of this gene shares sequence homology with
25 metalloprotease and also with thrombospondin, which is thought to be important in the activation of proteins and the processes of thrombopoiesis and metabolism.

This gene is expressed in many tissues, but especially in bladder, kidney, and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as
30 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of thrombopenia, hypertension, and other blood disfunctions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of
35 the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., urogenital, and reproductive tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID 5 NO. 174 as residues: Gly-8 to Leu-14, Met-18 to Phe-30.

The tissue distribution and homology to thrombospondin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of a variety of blood-related diseases.

10 FEATURES OF PROTEIN ENCODED BY GENE NO: 65

This gene is expressed primarily in tonsil, placenta, and fetal tissues. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of many diseases of the immune system. Similarly, 15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune and developmental tissues, and cancerous and wounded tissues) or bodily 20 fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or amniotic fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides 25 corresponding to this gene are useful for diagnosis and treatment of diseases of the immune system including many cancers such as lymphomas, leukemias, lymphocytomas, and the like.

FEATURES OF PROTEIN ENCODED BY GENE NO: 66

30 Polypeptides encoded by this gene share reasonable homology to steroid/thyroid hormone orphan nuclear receptor and to several additional orphan nuclear receptors isolated from several different tissues.

This gene is expressed primarily in testis. Therefore, polynucleotides and polypeptides of the invention are useful as 35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of testicular tumors, impotence, and other reproductive disorders. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., male reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or seminal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of diseases in the male reproductive system such as tumors of the testis and other reproductive disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 67

15 Polypeptides encoded by polynucleotides comprising this gene have a high degree of sequence identity with CTGF-4.

In one embodiment, the polypeptides of the invention comprise the sequence: MDSMPEPASRCLLPLLLLLLPAPELGPSQAGAEENDWVRLPSK CEVCKYVAVELKVVKPLRKQRQDTEVIGTVYGILDQKASGVKYTKSDLRLIEVTET

20 ICKRLLDYSLHKERTGSXRFAKGMSSETFETLHXLVHKGVKVMDIPYELWNE TSAEVADLKKQCDVLVEEFEEVIEDWYRNHQEEDLTEFLCANHVLKGKDTSC AEQWSGKKGDTAALGGKKSKKSIRAKAAGGRSSSSKQRKELGGLEGDPSP EEDEGIQKASPLTHSPPDEL(SEQ ID NO:300). Polynucleotides encoding these polypeptide sequences are also encompassed by the invention.

25 This gene is expressed in many tissues especially including cells in the immune system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for the diagnosis of cancers, immunological disorders, and neural diseases (such as spinocerebellar ataxia, bipolar affective disorder, schizophrenia, and autism), and other diseases featuring anticipation, neurodegeneration, or abnormalities of neurodevelopment. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nerve system, immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune cells and/or tissue, and cancerous and wounded tissues) or bodily

fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those 5 comprising a sequence shown in SEQ ID NO. 177 as residues: Ser-3 to Ser-9, Gly-36 to Val-43, Leu-45 to Gly-51.

FEATURES OF PROTEIN ENCODED BY GENE NO: 68

Polypeptides encoded by polynucleotides comprising this gene contain a zinc 10 finger homology domain. Such motifs are believed to be important for protein interactions, particularly with regard to gene regulation.

This gene is expressed primarily in T cells and the colon and, to a lesser extent, in the testes and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as 15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of many immune and digestive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the 20 immune and digestive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune, gastrointestinal, and reproductive system tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, or seminal fluid) or another tissue or cell sample taken from an individual having such a disorder, 25 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 178 as residues: Pro-12 to Lys-33, Asn-41 to His-46, Pro-48 to Ser-58, Gly-71 to Asp-78, Ala-94 to Gly-102, Ser-133 to Ser-140, Arg-197 to Lys-202.

30 The expression of this gene in T-cells indicates a potential role in the treatment and detection of immune disorders such as arthritis, asthma, immune deficiency diseases (such as AIDS), and leukemia. Expression of this gene in the colon indicates a potential role in the treatment and detection of colon disorders such as ulcers and colon cancer in addition to digestive disorders in general.

FEATURES OF PROTEIN ENCODED BY GENE NO: 69

The translation product of this gene shares sequence homology with neuroendocrine protein which is thought to be important in neuronal development and differentiation. A preferred embodiment of this gene comprises the following amino acid sequence: 5 MDGQKKNWKDKVVDLLYWRDIKKTGVVFGASLFLLSLTVF SIVSVTAYIALALLSVTISFRIYKGVIQAIQKSDEGHPFRAYLESEVAISEELVQKY SNSALGHVNCTIKELRRFLVDDLVDLSKFAVLMWVFTYVGALFNGLTLILAL ISLFSVPVITYERHQAQIDHYLGLANKNVKDAMAKIQAKIPGLKRKAE (SEQ ID NO:301). Particularly preferred are polynucleotides comprising polynucleotides 10 encoding this polypeptide sequence.

This gene is expressed in many different tissues, but primarily in brain, and, to a lesser extent, in fetal tissue, placenta, bone marrow, and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for diagnosis of neurodegenerative diseases and developmental disorders. 15 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system and during development, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neural, 20 developmental, and hemopoietic cells and tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those 25 comprising a sequence shown in SEQ ID NO. 179 as residues: Gln-47 to Gly-52, Leu-169 to Glu-174.

The predominant tissue distribution in brain and homology to neuroendocrine protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of neurodegenerative diseases and behavioral 30 disorders such as Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, schizophrenia, mania, dementia, paranoia, obsessive-compulsive disorder and panic disorder.

FEATURES OF PROTEIN ENCODED BY GENE NO: 70

35 Polypeptides encoded by polynucleotides comprising this gene share sequence identity with human hepatoma-derived growth factor (WPI 95-069304/10). As such, polynucleotides comprising this gene can be used for the recombinant production of the

protein, which can be used to encourage the growth of various animal cells, and for the purification of receptors. Additional embodiments of the invention comprise the following polypeptide sequences: MAVTLSLLLGGGRVCA (SEQ ID NO:302); PSLAVGSRPGGW RAQALLAGSRTPIPTGSRRNGSCRRWRAP (SEQ ID NO:303); and/or MAVTLSLLLGGRVCAPSLAVGSRPGGWRAQALLAGSRTPIPTGSRRNGSCRRWRAP (SEQ ID NO:304). Also contemplated are polynucleotides comprising polynucleotides encoding the aforementioned polypeptide sequences.

5 This gene is expressed primarily in brain and to a lesser extent in endothelium, T-cell, and tumors.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of many neurodegenerative diseases (for example, Alzheimer's Disease, ALS, and the like) and cancers (including, but not limited to neuroblastoma, glioblastoma, Schwannoma, astrocytoma, and the like). Similarly, 15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neural, and haematopoietic cells and tissue, and 20 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid or lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID 25 NO. 180 as residues: Pro-4 to Thr-10, Glu-25 to Trp-30, Leu-58 to Leu-69, Arg-82 to Thr-87, Ala-108 to His-115, Ser-124 to Glu-146, Pro-159 to Gly-176, Ser-182 to Glu-187, Leu-189 to Ser-198, Phe-208 to Asn-214.

25 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of many 30 neurodegenerative diseases and cancers.

FEATURES OF PROTEIN ENCODED BY GENE NO: 71

The translation product of this gene shares sequence homology with acrosin, trypsin, as well as trypsinogen precursor which are thought to be important in cell-cell 35 recognition and proteinase activity for protein cleavage and degradation. Preferred polynucleotide fragments comprise the following sequence:
GATGTTACACAGCTCTTAATAATAGTGGCCATAGCTGTAATAACAATGACA

ACAGTAGGTAACGGTAGTCATACCAACAGTAGGGCAGTGCATTTATATTAC
AACTGGTTCTTGCTCTAGTAGGCTGGGATGGGTAAAGACGGACAGGGC
TGGCGCAGACCCTTCCITCTCCTCTCCAGCCCACAGTGATCTGGCTTTA
CAGACAGCCTGCTTCCATTCACTAGTAGTGTGGAAAGTCCCTCTGGCTTAGC
5 AATACCCCTGAGACCTTGTCACTGGGCTGTGCTCTCCCTGGGATGCTGG
GAGCACCAAGTGTGGCCGAGCTAGGGCTGCTGACTTCCTCTGGCGCCTCT
GGGCTGCGAGGGTCTTATAGGAATTGAGGCCCTTGCTGCTCCAAGAAA
TGCAGGGCTGTGGGCARAGGGKTGTACCCAAGGGACTCTTGCTCTGTGT
CTGACTTTGGGGRATCC (SEQ ID NO:305); CACAGCTTTAATAATAGTGGC
10 CATAGCTGTAATAACAATGACA ACAGTAGGTAACG (SEQ ID NO:306);
TGTGTCTCTCCCTGGGATGCTGGGAGCACCAAGTGTGGCCGAGCTAGGGCT
GCTGACTT (SEQ ID NO:307); GCGAGGGTCTTATAGGAATTGAGGCCCTT
TGCTGCTCCAAGAAATGCTGAGGCTGTGGGCARAGGGKTGTACCCAAGGG
GACT (SEQ ID NO:308). Also preferred are polypeptide fragments encoded by these
15 polynucleotide fragments.

This gene is expressed primarily in cheek carcinoma and to a lesser extent in uterine and pancreatic cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cheek cancers or cancers of uterine and pancreatic origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neoplastic tissues, 25 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., epithelial, endocrine, and reproductive tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and saliva) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., 30 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to acrosin and trypsin indicates that 35 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of cancers. The homology to acrosin and trypsin may indicate the gene function in tumor metastasis or migration since in both cases cell-cell interaction and extracellular matrix degradation may be involved. The gene product can also be used as a target for cancer immunotherapy or as a diagnostic marker.

FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene is expressed primarily in T helper cells I, T-cells stimulated with PHA for 24 hours, and in a placenta Nb2HP cDNA library.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of many immunodeficiencies and disorders (especially autoimmune diseases). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

10 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune, and haematopoietic cells and tissue, and cancerous and wounded tissue) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid and

15 lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of autoimmune diseases, immunodeficiencies, and other immune system disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 73

This gene is expressed primarily in 7 week old early stage human, human chronic synovitis, and infant brain.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of chronic synovitis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

30 of the above tissues or cells, particularly of the synovium, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., developmental, differentiating, and neural tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and amniotic fluid) or another tissue or cell sample taken from an individual

35 having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 183 as residues: Ser-44 to Pro-49.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of chronic 5 synovitis and other disorders of the synovium.

FEATURES OF PROTEIN ENCODED BY GENE NO: 74

Polypeptides encoded by polynucleotides comprising this gene exhibit sequence homology to a number of mucin-like extracellular or cell surface proteins. In one 10 embodiment polypeptides of the invention comprise the following sequence:

MVGPVTLHKKIHTTTVLFIVQIHLILLQAITQAK (SEQ ID NO:309); LQMHLMLQ MTGLSILALLGKSTTTIVEQKFHNGKNQKSGLKENRDKKKQTRWQSTASQKI GITEER (SEQ ID NO:310); and/or MVGPVTLHKKIHTTTVLFIVQIHLILLQAITQ AKLQMHLMLQMTGLSILALLGKSTTTIVEQKFHNGKNQKSGLKENRDKKKQ 15 TRWQSTASQKIGITEER (SEQ ID NO:311). Polynucleotides encoding the aforementioned polypeptides are also contemplated embodiments of the invention.

This gene is expressed primarily in ovarian cancer, endometrial tumor, B-cell lymphoma, brain-medulloblastoma, hepatocellular tumor, osteosarcoma, and T- and B-cells.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, Ovarian cancer, endometrial tumor, B-cell lymphoma, brain medulloblastoma, hepatocellular tumor, and osteosarcoma. Similarly, polypeptides and 25 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, bone, T-cells and other 30 cells of the immune system, and B cells and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid and lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred 35 epitopes include those comprising a sequence shown in SEQ ID NO. 184 as residues: Met-1 to Lys-12, Leu-14 to Asn-35, Arg-42 to Asn-58, Ser-65 to Trp-90, Ser-95 to Asn-129, Phe-136 to Arg-144, Met-159 to Ala-167, Thr-179 to Tyr-187, Pro-190 to

Val-201, Gln-226 to Phe-235, Pro-254 to His-272, Thr-288 to Thr-293, Thr-383 to Ser-391, Asp-398 to Tyr-405, Ile-410 to Asn-416, Ala-449 to Lys-458.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of ovarian cancer, 5 endometrial tumors, B-cell lymphoma, brain medulloblastoma, hepatocellular tumor, and osteosarcoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 75

An additional preferred polypeptide sequence derived from the polynucleotide of 10 this contig comprises the following amino acid sequence: MQTCPLVGTLLTRNMDG YTCAVVTSTSFWIISAWXLWKGSPSTSMPMPETPLRTLCCDKMPSSIFSSLMTD GRA (SEQ ID NO:312). Polynucleotides encoding these polypeptides are also provided. This polypeptide sequence has sequence homology with a *Drosophila melanogaster* male germ-line specific transcript which encodes a putative protamine 15 molecule (see, gil608696).

This gene is expressed primarily in breast tissue and to a lesser extent in various other fetal and adult cells and tissues, especially those comprising endocrine organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental and reproductive defects. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression 25 of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., breast and/or other ductile secretory tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and milk) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level 30 in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of developmental, reproductive and growth and metabolic disorders.

35 FEATURES OF PROTEIN ENCODED BY GENE NO: 76

In one embodiment, the polypeptides of the invention comprise the sequence:
MTLIQNCWYSWLFFGFFFHFLRKSISIFSIFLVCFRILALGPTCFLVWFWKAFFR

HILIFICLSREVFRPRCFLVYFR (SEQ ID NO:313). This polypeptide sequence has sequence homology with the MURF4 protein of *Herpetomonas muscarum* (S43288). Such RNA-editing enzymes may be useful as molecular targets in the intervention of the life cycle of trypanosomes and other protozoa. Polynucleotides encoding these 5 polypeptides are also encompassed by the invention.

This gene is expressed primarily in fetal liver and spleen, osteosarcoma and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 10 biological sample and for diagnosis of liver tumors, osteosarcoma, and other cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be 15 routinely detected in certain tissues and cell types (e.g., hepatic, developmental, and differentiating tissue, bone cells, liver and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily 20 fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of cancers such as liver tumor and osteosarcoma.

25 FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This gene is expressed primarily in T cell lymphoma and monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a 30 biological sample and for diagnosis of T-cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., immune and hematopoietic cells and tissues, and cancerous and 35 wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in

healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 187 as residues: Thr-1 to Ser-9.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of T-cell lymphoma.

FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed primarily in tonsils and a bone marrow cell line. Therefore, polynucleotides and polypeptides of the invention are useful as 10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, 15 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., haematopoietic and immune cells and tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the 20 expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immunological disorders.

25

FEATURES OF PROTEIN ENCODED BY GENE NO: 79

In one embodiment, the polypeptides of the invention comprise the sequence: MGTRAQVTGRLPIPPPAPGLPFSAXEPLQGQLRRVSSSRGGFPGLALQLLRSE TVKAYVNNEINILASF (SEQ ID NO:314) and/or MLVRTRPSQPLPLPGVGLGGP 30 RSGDPPESTELRKGPGLFA (SEQ ID NO:315). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in brain, placenta, bone marrow, keratinocyte, fetal liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as 35 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of brain and skin related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing

immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and skin system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neural, reproductive, and hepatic tissues, 5 keratinocytes, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a 10 sequence shown in SEQ ID NO. 189 as residues: Phe-13 to Leu-18.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of many brain and skin related diseases.

15 FEATURES OF PROTEIN ENCODED BY GENE NO: 80

The translation product of this gene shares sequence homology with mouse RNA Polymerase I which is thought to be important in gene transcription process.

This gene is expressed primarily in HEL cell line and aorta endothelial cells and to a lesser extent in Jurkat T-cells.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis and treatment of cancer and autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell 25 type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial, haematopoietic tissues, cardiovascular tissue, and T-cells and other cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial 30 fluid, spinal fluid, and lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 190 as residues: Lys-25 to Arg-32.

35 The tissue distribution and homology to mouse RNA polymerase I indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of immune diseases and cardiovascular diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 81

In one embodiment, the polypeptides of the invention comprise the sequence:

5 MCPVCGRALSSPGSLGRHLLIHSEDQRSNCAVCGARFTSHATFNSEKLPEVLN
MESLPTVHNEGPSSAEGKDIASFPPVYPAGILLVCNNCAA~~YRKXLEAQTPSVX~~
KWALRRQNEPLEVRLQRLERERTAKKSRRDNETPEEREVRMRDREAKRLQR
MQETDEQRARRLQRDREAMRLKRANETPEKRQARLIREREAKRLKRRLEKMD
10 MMLRAQFGQDPSAMAALAAEMNFFQLPVGVELDXQLLGKMAFEEQNSSLH
(SEQ ID NO:316). This polypeptide shares sequence homology with human trichohyalin
which is thought to be important in gene regulation. Polynucleotides encoding this
polypeptide are also encompassed by the invention.

This gene is expressed primarily in brain tissue and to a lesser extent in
apoptotic T-cell and B-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as
15 reagents for differential identification of the tissue(s) or cell type(s) present in a
biological sample and for diagnosis and treatment of growth disorders,
neurodegenerative diseases, and endocrine disorders. Similarly, polypeptides and
antibodies directed to these polypeptides are useful in providing immunological probes
20 for differential identification of the tissue(s) or cell type(s). For a number of disorders
of the above tissues or cells, particularly of the neural and immune systems, expression
of this gene at significantly higher or lower levels may be routinely detected in certain
tissues and cell types (e.g., neural tissues, T-cells, B-cells and other cells and tissue of
the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum,
25 plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from
an individual having such a disorder, relative to the standard gene expression level, i.e.,
the expression level in healthy tissue or bodily fluid from an individual not having the
disorder.

The tissue distribution and homology to DNA binding protein indicates that
polynucleotides and polypeptides corresponding to this gene are useful for the
30 diagnosis and treatment of immune and neurological diseases.

FEATURES OF PROTEIN ENCODED BY GENE NO: 82

In one embodiment, the polypeptides of the invention comprise the sequence:

MDHSHHMGMSYMDSNSTMQPSHHPTTSASHSHGGGDSSMMMPMTFYFG
35 FKNVELLFSGLVINTAGEMAGAFVAVFLLAMFYEGLKIARESLLRKSQVSIRYN
SMPVPGPNGTILMETHKTVGQQMLSFPHLLQTVLHIIQVVISYFLMLIFMTYNG
YLCIAAXAAGAGTGYFLFSWKKAVVVDITEHCH (SEQ ID NO:317). This

polypeptide is thought to function in mediating the uptake of copper and other metal ions by cells. Polynucleotides encoding this polypeptide are also encompassed by the invention.

5 This gene is expressed primarily in osteosarcoma and to a lesser extent in T-cell and bone marrow stromal cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for treatment and diagnosis of osteosarcoma and copper and other metal uptake disorders. Similarly, polypeptides and antibodies directed to these 10 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic tissue and cancerous and wounded tissues) or bodily fluids (e.g., 15 serum, plasma, urine, synovial fluid, spinal fluid, and lymph) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 192 as residues: Ser-24 to Ser-29.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the prevention or treatment of osteosarcoma and copper or other metal uptake disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 83

25 This gene is expressed primarily in skin tumor and to a lesser extent in apoptic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 30 not limited to, skin tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., epithelial and 35 hematopoietic tissues, and T-cells and other tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, and spinal fluid) or another tissue or cell sample taken from an individual having

such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 193 as residues: Leu-51 to Gly-77, Ile-117 to Pro-125.

5 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis the treatment of skin tumor.

FEATURES OF PROTEIN ENCODED BY GENE NO: 84

This gene is expressed primarily in testis.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, infertility and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

15 for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and seminal fluid) or

20 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of reproductive disease and endocrine disorders.

FEATURES OF PROTEIN ENCODED BY GENE NO: 85

In one embodiment, the polypeptides of the invention comprise the sequence:

30 MVQPCGACAKTXWKACSSCCSSPCCCLQERWPXPXAXCPEXGPSSHPGIQALC
AVAVVYLSPSSRLDWSLAPLFVPSLAAGETPLTQPAWALTNTLGHGQPAQDR
LPALGHCAPISVLGLGSS (SEQ ID NO:318). Polynucleotides encoding this polypeptide sequence are also encompassed by the invention.

This gene is expressed primarily in kidney cortex, frontal cortex, spinal cord and hippocampus.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, kidney fibrosis, schizophrenia and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system,

5 expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial, neural and endocrine tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level
10 in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 195 as residues: Cys-27 to Tyr-33, Thr-38 to Gly-43, Leu-125 to Gly-130.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of neurological disorders and
15 kidney diseases..

FEATURES OF PROTEIN ENCODED BY GENE NO: 86

This gene is expressed primarily in resting T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as
20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T-cell related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above
25 tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic and immune cells and tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid, spinal fluid, and lymph) or another tissue or cell sample taken from an individual having such a disorder,
30 relative to the standard gene expression level, (i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder). Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 196 as residues: Thr-54 to Ile-59.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of immune diseases.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep.	First AA SEQ ID NO: Y	Last AA of Sig Pep.	First AA Secreted Portion	Last AA of AA of ORF
1	HOAAE80	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	11 1220	264	1220	288	288	111	1 26 27 31
2	HODDN92	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	12	1939	294	1939	434	112	1 26 27 35
3	HOSB96	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	13	2602	672	1811	690	113	1 30 31 219
4	HOVA158	209012 04/28/97 209089 06/05/97	pSport1	14	808	1	808	28	114	1 26 27 31
5	HPBDD36	209012 04/28/97 209089 06/05/97	pBluescript SK-	15	864	87	831	147	115	1 18 19 26
6	HPDDC77	209012 04/28/97 209089 06/05/97	pBluescript SK-	16	2361	455	1442	510	116	1 29 30 131
7	HPEBD85	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	17	803	1	803	81	117	1 20 21 64

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
		04/28/97 209089 06/05/97	04/28/97 209089 06/05/97	Uni-ZAP XR	18	1794	1051	1757	578	118	1
8	HPFCX38	209012	04/28/97 209089 06/05/97	Uni-ZAP XR	19	1037	1	1037	467	119	1
9	HPFCY51	209012	04/28/97 209089 06/05/97	Uni-ZAP XR	97	1052	1	1052	30	30	31
9	HPFCY51	209012	04/28/97 209089 06/05/97	Uni-ZAP XR	97	1052	1	1052	30	197	1
10	HPMGQ80	209012	04/28/97 209089 06/05/97	Uni-ZAP XR	20	1309	157	1309	360	360	120
11	HPRITG55	209012	pBluescript	21	1081	55	1014	237	237	121	1
12	HROAN56	209012	04/28/97 209089 06/05/97	Uni-ZAP XR	22	807	1	807	26	26	122

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	NT Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Start Signal Codon	NT SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion	AA of First AA of Secreted Portion	Last AA of First AA of Secreted Portion
13	HSABI42	209012 04/28/97 209089 06/05/97	pBluescript SK-	23	632	1	596	190	190	123	1	15	16	21
14	HSAUW44	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	24	1358	1	1358	372	372	124	1	30	31	34
15	HSDES04	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	25	1376	686	1376	146	146	125	1	33	34	318
15	HSDES04	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	98	929	57	929	291	291	198	1	28	29	61
16	HSHBQ68	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	26	2923	195	2642	211	211	126	1	23	24	58
17	HSKBO20	209012 04/28/97 209089 06/05/97	Uni-ZAP XR	27	775	1	501	308	127	1	28	29	98	
18	HSKNM85	209012 04/28/97 209089	pBluescript	28	534	1	534	122	122	1	19	20	28	

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ	NT ID	5' NT of Total NT Seq: X	3' NT of Clone Seq.	5' NT of Total NT Seq: X	NT SEQ	5' NT of Total NT Seq: X	AA ID	First AA of Signal Seq. Codon	Last AA of Signal Seq. Codon	First AA of Secreted Portion	Last AA of Secreted Portion	First AA of ORF	Last AA of ORF
27	HT4AI54	209090 06/05/97	Uni-ZAP XR	37	2279	1387	2279	29	29	137	1	24	25	25	288	
27	HT4AI54	209090 06/05/97	Uni-ZAP XR	100	952	1	952	199	199	200	1					10
28	HTEH93	209090 06/05/97	Uni-ZAP XR	38	745	1	745	187	187	138	1	24	25	25	25	113
29	HTGCQ82	209090 06/05/97	Uni-ZAP XR	39	1718	70	1718	114	114	139	1	23	24	24	24	119
30	HTLAB25	209090 06/05/97	Uni-ZAP XR	40	1966	321	1966	449	449	140	1	1	1	1	2	438
31	HTLAV68	209090 06/05/97	Uni-ZAP XR	41	972	1	972	78	78	141	1	35	36	36	36	162
32	HTLDQ11	209090 06/05/97	Uni-ZAP XR	42	1536	1	1536	213	213	142	1	36	37	37	37	72
33	HTOBX52	209090 06/05/97	Uni-ZAP XR	43	2541	1743	2541	3	143	1	4	5	5	5	5	123
34	HTTCN24	209090 06/05/97	Uni-ZAP XR	44	2418	918	2290	188	188	144	1	30	31	31	31	138
34	HTTCN24	209090 06/05/97	Uni-ZAP XR	101	1545	123	1545	345	345	201	1	39	40	40	40	50
35	HTXCS21	209090 06/05/97	Uni-ZAP XR	45	1337	657	1309	76	76	145	1	24	25	25	25	356
35	HTXCS21	209090 06/05/97	Uni-ZAP XR	102	1322	641	1293		1203	202	1					13
36	HUFAC49	209090 06/05/97	pSport1	46	1276	1	1276	105	105	146	1	17	18	18	18	39

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID: NO: X	NT Total Seq. NO: X	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA SEQ ID: NO: Y	First AA of Signal Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
37	HAIDK60	209090 06/05/97	Uni-ZAP XR	47	1282	1	1282	528	147	1	30
37	HAIDK60	209090 06/05/97	Uni-ZAP XR	103	276	1	276	14	203	1	25
38	HARAG28	209090 06/05/97	pBluescript SK-	48	645	1	645	150	148	1	16
38	HARAG28	209090 06/05/97	pBluescript SK-	104	381	1	381	154	204	1	18
39	HBMBB80	209090 06/05/97	pBluescript	49	1495	2	1495	23	149	1	30
39	HBMBB80	209090 06/05/97	pBluescript	105	638	1	638	196	205	1	16
40	HCEGR33	209090 06/05/97	Uni-ZAP XR	50	1630	1	1630	243	243	150	1
41	HSXBP68	209090 06/05/97	Uni-ZAP XR	51	2420	1009	2252	79	79	151	1
41	HSXBP68	209090 06/05/97	Uni-ZAP XR	106	2246	835	2079	985	985	206	1
42	HFFAT33	209090 06/05/97	Lambda ZAP II	52	1172	166	802	209	209	152	1
43	HFGAG96	209090 06/05/97	Uni-ZAP XR	53	1589	885	1446	189	189	153	1
43	HFGAG96	209090 06/05/97	Uni-ZAP XR	107	1105	1	1105		247	207	1
44	HETFJ05	209076 05/22/97	Uni-ZAP XR	54	2074	1	2065	75	75	154	1

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	AA SEQ ID of Signal Pep Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
45	HLTEY63	209076 05/22/97	Uni-ZAP XR	55	1483	1	1280	86	155	1	18
46	HMSJU68	209076 05/22/97	Uni-ZAP XR	56	1123	4	1123	272	272	1	31
47	HOSCZ41	209076 05/22/97	Uni-ZAP XR	57	1239	117	1222	178	178	1	20
48	HSHAV28	209076 05/22/97	Uni-ZAP XR	58	803	105	719	378	378	1	21
49	HSQEA85	209076 05/22/97	Uni-ZAP XR	59	995	1	995	98	98	159	1
50	HSTAG52	209076 05/22/97	Uni-ZAP XR	60	966	114	966	191	191	160	1
51	HBNAJ22	209076 05/22/97	Uni-ZAP XR	61	262	1	262	28	28	161	1
52	HBXGP76	209076 05/22/97	ZAP Express	62	753	1	753	34	34	162	1
53	HE6GL64	209076 05/22/97	Uni-ZAP XR	63	739	1	739	132	132	163	1
54	HESAL35	209076 05/22/97	Uni-ZAP XR	64	476	1	476	20	20	164	1
55	HETBB70	209076 05/22/97	Uni-ZAP XR	65	754	14	754	263	263	1	17
56	HLHAY19	209076 05/22/97	Uni-ZAP XR	66	1890	8	1890	18	18	166	1
57	HLTER45	209076 05/22/97	Uni-ZAP XR	67	1614	557	1614	578	578	167	1

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Total NT Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Start Codon	5' NT of AA of Signal Start Codon	First AA SEQ ID NO: Y	Last AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA Secreted Portion	Last AA Secreted Portion	First AA of Sig Pep	Last AA of Sig Pep	First AA Secreted Portion	Last AA Secreted Portion
58	HNHAL34	209076 05/22/97	Uni-ZAP XR	68	596	1	596	90	90	168	1	18	19	18	19	39	39
59	HOSFF78	209076 05/22/97	Uni-ZAP XR	69	1524	791	1524	846	846	169	1	34	35	35	35	46	46
60	HSKDV92	209076 05/22/97	Uni-ZAP XR	70	819	53	819			158	170	1	32	33	33	33	33
61	HFCCU63	209076 05/22/97	Uni-ZAP XR	71	1442	1	1442	12	12	171	1					4	4
62	HLTCS34	209076 05/22/97	Uni-ZAP XR	72	1223	1	1223	227	227	172	1	17	18	18	18	24	24
63	HPMCC16	209086 05/29/97	Uni-ZAP XR	73	1814	1024	1814	85	85	173	1	19	20	19	20	262	262
64	HOUQC17	209086 05/29/97	Uni-ZAP XR	74	4712	1	4693	508	508	174	1	51	52	51	52	967	967
65	HTDAG66	209086 05/29/97	pSport1	75	1885	262	1885	369	369	175	1					18	18
66	HTLBC79	209086 05/29/97	Uni-ZAP XR	76	890	1	890	17	17	176	1	1	2	1	2	205	205
67	HTOFC34	209086 05/29/97	Uni-ZAP XR	77	1657	356	1645	434	434	177	1	31	32	31	32	54	54
68	H2CBJ08	209086 05/29/97	pBluescript SK-	78	2015	13	2015	70	70	178	1	17	18	17	18	435	435
69	HAGFT48	209086 05/29/97	Uni-ZAP XR	79	1213	242	1213			290	179	1	23	24	24	174	174
70	HCE5M29	209086 05/29/97	Uni-ZAP XR	80	1391	23	1353	251	251	180	1	1	2	1	2	219	219

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID	NT Total NO: X	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA SEQ	5' NT of AA ID	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of Secreted Portion
71	HTPBQ83	209076 05/22/97	Uni-ZAP XR	81	1008	146	1008	431	181	1		5
72	HCFNN01	209086 05/29/97	pSport I	82	1261	154	1261	254	182	1	27	28
73	HE7TF86	209086 05/29/97	Uni-ZAP XR	83	1045	241	986	426	183	1	23	24
74	HGBAC11	209086 05/29/97	Uni-ZAP XR	84	2877	1	2272	85	85	184	1	2
75	HHGAU81	209086 05/29/97 II	Lambda ZAP	85	1367	747	1367	323	323	185	1	25
76	HLCAA05	209086 05/29/97	Uni-ZAP XR	86	1009	1	1009	276	276	186	1	
77	HMSCD68	209086 05/29/97	Uni-ZAP XR	87	1367	1	1367	254	187	1		
78	HMWDZ81	209086 05/29/97	Uni-Zap XR	88	1088	1	883	214	214	188	1	22
79	HMWGQ73	209086 05/29/97	Uni-Zap XR	89	1861	875	1861	1160	189	1	15	16
80	HOECN31	209086 05/29/97	Uni-ZAP XR	90	1259	34	1259	338	338	190	1	28
81	HPTRF90	209086 05/29/97	pBluescript	91	1566	450	1552	593	593	191	1	28
82	HSRDFH01	209086 05/29/97	Uni-ZAP XR	92	1593	107	1593	379	379	192	1	22
83	HSAWD74	209126 06/19/97	Uni-ZAP XR	93	970	106	970	142	142	193	1	26

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	NT SEQ ID NO: X	5' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of AA ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
83	HSTBE27	209086 05/29/97	Uni-ZAP XR	110	646	117	646	122	210	1
84	HTEJO12	209086 05/29/97	Uni-ZAP XR	94	934	1	934	202	194	1
85	HTLAB43	209086 05/29/97	Uni-ZAP XR	95	1392	199	1392	384	195	1
86	HTWCT03	209086 05/29/97	pSport1	96	1963	1	1963	334	196	1
								26	27	101

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The 5 overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain 10 multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT 15 of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified 20 as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted 25 first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and 30 otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic 35 methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

5 The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40 (1988).
10 Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the 15 cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 20 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra*.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide 25 was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results 30 shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence 35 shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

10 Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

15 By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other 20 words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF 25 (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between 30 a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result 35 of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization

Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query

amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions,

5 interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%; 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be

10 determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and
15 subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window
20 Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity.

25 For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of
30 the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are
35 considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired 5 residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. 10 This time the deletions are internal deletions so there are no residues at the N- or C- termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query 15 sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or 20 activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety 25 of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. 30 Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be 35 deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after

deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

Moreover, ample evidence demonstrates that variants often retain a biological 5 activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible 10 amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

15 Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form 20 are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show 25 substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main 30 strategies for studying the tolerance of an amino acid sequence to change.

The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions 35 where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham 5 and Wells, *Science* 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the 10 protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues 15 Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, 20 where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino 25 acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins 30 with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., *Clin. Exp. Immunol.* 2:331-340 (1967); Robbins et al., *Diabetes* 36: 838-845 (1987); Cleland et al., *Crit. Rev. Therapeutic Drug Carrier Systems* 10:307-377 (1993).)

Polynucleotide and Polypeptide Fragments

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 5 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers 10 as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-15 450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the 20 deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 30 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

35 Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the

carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any 5 combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Particularly, N-terminal deletions of the polypeptide of the present invention can be described by the general formula m-p, where p is the total number of amino acids in 10 the polypeptide and m is an integer from 2 to (p-1), and where both of these integers (m & p) correspond to the position of the amino acid residue identified in SEQ ID NO:Y.

Moreover, C-terminal deletions of the polypeptide of the present invention can also be described by the general formula 1-n, where n is an integer from 2 to (p-1), and again where these integers (n & p) correspond to the position of the amino acid residue 15 identified in SEQ ID NO:Y.

The invention also provides polypeptides having one or more amino acids deleted from both the amino and the carboxyl termini, which may be described generally as having residues m-n of SEQ ID NO:Y, where m and n are integers as described above.

20 Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

25 Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

30 Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

Epitopes & Antibodies

35 In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an

epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 5 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at 10 least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to 15 methods well known in the art. (See, for instance, Sutcliffe et al., *supra*; Wilson et al., *supra*; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if 20 it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is 25 meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')2 fragments) which are capable of specifically binding to protein. Fab and F(ab')2 fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, 30 as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

Fusion Proteins

35 Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the

polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

5 Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

10 Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final 15 preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

20 Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of 25 mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

30 Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the 35 fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D.

Bennett et al., *J. Molecular Recognition* 8:52-58 (1995); K. Johanson et al., *J. Biol. Chem.* 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In 5 preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. 10 Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., *Cell* 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

15 Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the 20 latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then 25 transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The 30 expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

35 As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance

genes for culturing in *E. coli* and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells; insect cells such as *Drosophila S2* and *Spodoptera Sf9* cells; animal cells such as CHO, COS, 5 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, 10 pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium 15 phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

20 A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most 25 preferably, high performance liquid chromatography ("HPLC") is employed for purification.

30 Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or 35 eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the 35 translation initiation codon generally is removed with high efficiency from any protein

after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

5

Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

10

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

15

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing 20 the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

25

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

30

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence *in situ* hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

35

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are

more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage 5 analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, *Mendelian Inheritance in Man* (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease 10 could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or 15 translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the 20 mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected 25 individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred 30 polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., *Nucl. Acids Res.* 6:3073 (1979); Cooney et al., *Science* 241:456 (1988); and Dervan et al., *Science* 251:1360 (1991)) or to the mRNA itself (antisense - Okano, *J. Neurochem.* 56:560 (1991); *Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression*, CRC 35 Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model

systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of

unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for

5 contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers

10 for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The

15 following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-20 3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and 25 technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected *in vivo* by imaging. Antibody labels or markers for *in vivo* imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

35 A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic

resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human 5 subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of ^{99m}Tc . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The 10 Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene 15 expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to 20 supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired 25 response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such 30 as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a 35 recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and 5 polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

Immune Activity

10 A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells 15 from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

20 A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic 25 cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency 30 (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood 35 coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can

decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in 5 treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, 10 differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, 15 antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuropathy, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune 20 inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

25 A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The 30 administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

35 Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic

shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or 5 IL-1.)

Hyperproliferative Disorders

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present 10 invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing 15 antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

20 Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, 25 pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a 30 polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstron's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

Infectious Disease

35 A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases

may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

5 Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, 10 Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., 15 Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, 20 Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that 25 can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Nocardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, 30 Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Menigococcal), Pasteurellacea Infections (e.g., Actinobacillus, Heamophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, 35 and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS

related infections), paronychia, prosthetic-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria,

5 Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocytoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

10 Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas.

15 These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or

20 diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (*ex vivo* therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

Regeneration

A polynucleotide or polypeptide of the present invention can be used to

30 differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal

or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase 5 regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue 10 regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate 15 nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, 20 Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

Chemotaxis

25 A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular 30 trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. 35 For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

5

Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit 10 (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural 15 or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

20 Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed 25 polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a 30 labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

35 Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

5 All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

10 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with 15 a polypeptide of the invention, (b) assaying a biological activity , and (b) determining if a biological activity of the polypeptide has been altered.

Other Activities

20 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

25 A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

30 A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

35 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

5 Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the

10 10 Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

15 Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the

20 15 Clone Sequence as defined for SEQ ID NO:X in Table 1.

25 Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

30 A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under 5 stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which 10 comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide 15 sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at 20 least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

25 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

30 A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

35 A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method

comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

5 Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined
10 from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95%
15 identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

20 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

25 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous
30 nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

35 The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95%

identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide

comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone

5 identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

10 Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as

15 defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide

20 molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition

25 associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a

30 sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

35 In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected

5 from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

10 Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid

15 sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising

20 inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising

25 culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID

30 NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the

35 deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase 5 the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

10 Examples

Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. 15 Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector 20 "Lambda Zap," the corresponding deposited clone is in "pBluescript."

	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited Plasmid</u>
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
	Zap Express	pBK
25	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSport 2.0	pCMVSport 2.0
	pCMVSport 3.0	pCMVSport 3.0
	pCR®2.1	pCR®2.1
30	Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are	
35	commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1	

Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

5 Vectors pSport1, pCMVSport 2.0 and pCMVSport 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR®2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

10 20 The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

15 25

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

30 Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with ^{32}P - γ -ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).)

The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

10 Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25 μ l of reaction mixture with

15 0.5 ug of the above cDNA template. A convenient reaction mixture is 1.5-5 mM MgCl₂, 0.01% (w/v) gelatin, 20 μ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product

20 is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods

25 include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

30 Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to

35 generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then 5 be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA 10 synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

15

Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR 20 using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

Example 3: Tissue Distribution of Polypeptide

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, 25 among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P³² using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is 30 then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are 35 mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

Example 4: Chromosomal Mapping of the Polynucleotides

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This 5 primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on 10 either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

Example 5: Bacterial Expression of a Polypeptide

15 A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product 20 into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

25 The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the *E. coli* strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses 30 the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The 35 cells are grown to an optical density 600 (O.D.⁶⁰⁰) of between 0.4 and 0.6. IPTG

(Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by 5 centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrolo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high 10 affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 15 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 20 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM immidazole. Immidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

25 In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

30 DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA

insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

5 The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

Example 6: Purification of a Polypeptide from an Inclusion Body

10 The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit 15 weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

15 The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

20 The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C 25 overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 30 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area

(e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a 5 stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion 10 (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant A_{280} 15 monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 μ g of purified protein is loaded. 20 The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

Example 7: Cloning and Expression of a Polypeptide in a Baculovirus

Expression System

25 In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient 30 polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that 35 express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

15 The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

20 The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.).

25 The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

30 Five μ g of a plasmid containing the polynucleotide is co-transfected with 1.0 μ g of a commercially available linearized baculovirus DNA ("BaculoGoldTM baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One μ g of BaculoGoldTM virus DNA and 5 μ g of the plasmid are mixed in a sterile well of a microtiter plate containing 50 μ l of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 μ l Lipofectin plus 90 μ l Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm

tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

5 After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture 10 and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 15 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection 20 ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of ³⁵S-methionine and 5 µCi ³⁵S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins 25 in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

30 **Example 8: Expression of a Polypeptide in Mammalian Cells**

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates

the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved 5 with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, 10 for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSport 2.0, and pCMVSport 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

15 Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the 20 encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., *J. Biol. Chem.* 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., *Biochem. et Biophys. Acta*, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., *Biotechnology* 9:64-68 (1991).) Another useful selection marker is 25 the enzyme glutamine synthase (GS) (Murphy et al., *Biochem J.* 227:277-279 (1991); Bebbington et al., *Bio/Technology* 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a 30 chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., *Molecular and Cellular Biology*, 438-447 (March, 1985)) plus a fragment of the 35 CMV-enhancer (Boshart et al., *Cell* 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the

polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by 5 procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a 10 heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

15 The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for 20 transfection. Five μ g of the expression plasmid pC6 is cotransfected with 0.5 μ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are 25 trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of 30 methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1 μ M, 2 μ M, 5 μ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200 μ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., *Nature* 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGCCAAATCTTCTGACAAAATCACACATGCCACCGTGCC
CAGCACCTGAATTGAGGGTGCACCGTCAGTCCTCTCCCCCCTAAACCC
35 CAAGGACACCCCTCATGATCTCCGGACTCCTGAGGTACATGCGTGGTGG
GGACGTAAGCCACGAAGACCCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC

AGCACGTACCGTGTGGTCAGCGCCTCACCGCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCCACAGGT
GTACACCCCTGCCCTATCCCGGGATGAGCTGACCAAGAACCAAGGTCAACCT
5 GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCGTGCTGG
ACTCCGACGGCTCCTCTCCTCTACAGCAAGCTACCGTGGACAAGAGCA
GGTGGCAGCAGGGAACGTCTCTCATGCTCCGTGATGCATGAGGCTCTGC
ACAACCACACTACACCGAGAACAGGCCTCCCTGTCTCCGGTAAATGAGTGC
10 GACGGCCCGACTCTAGAGGAT (SEQ ID NO:1).

Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

20 In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., *Nature* 256:495 (1975); Köhler et al., *Eur. J. Immunol.* 6:511 (1976); Köhler et al., *Eur. J. Immunol.* 6:292 (1976); Hammerling et al., in: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at 30 about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as

described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be

5 produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells,

10 and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

15 It will be appreciated that Fab and F(ab')2 and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')2 fragments). Alternatively, secreted protein-binding fragments can be produced through the application of

20 recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art.

25 (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

30

Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in

35 Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a

working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The 5 PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2×10^5 cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x 10 Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in 15 Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of 20 transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off 25 PBS rinse, and person B, using a12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl₂ (anhyd); 0.00130 mg/L 30 CuSO₄·5H₂O; 0.050 mg/L of Fe(NO₃)₃·9H₂O; 0.417 mg/L of FeSO₄·7H₂O; 311.80 mg/L of KCl; 28.64 mg/L of MgCl₂; 48.84 mg/L of MgSO₄; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO₃; 62.50 mg/L of NaH₂PO₄·H₂O; 71.02 mg/L of Na₂HPO₄; .4320 mg/L of ZnSO₄·7H₂O; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic 35 Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitic Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of

Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L-Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H₂O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H₂O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H₂O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H₂O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B₁₂; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

Example 12: Construction of GAS Reporter Construct

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive

5 responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is

10 Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

15 The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

20 The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and

25 (b) Class 2 includes IFN- α , IFN- γ , and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

30 Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are

35 known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	<u>Ligand</u>	tyk2	<u>JAKs</u>			<u>STATS</u>	<u>GAS(elements) or ISRE</u>
			<u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>		
<u>IFN family</u>							
5	IFN-a/B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	IL-10	+	?	?	-	1,3	
<u>gp130 family</u>							
10	IL-6 (Pleiotropic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	IL-11 (Pleiotropic)	?	+	?	?	1,3	
	OnM (Pleiotropic)	?	+	+	?	1,3	
	LIF (Pleiotropic)	?	+	+	?	1,3	
15	CNTF (Pleiotropic)	-/+	+	+	?	1,3	
	G-CSF (Pleiotropic)	?	+	?	?	1,3	
	IL-12 (Pleiotropic)	+	-	+	+	1,3	
<u>g-C family</u>							
20	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >> Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
25	IL-15	?	+	?	+	5	GAS
<u>gp140 family</u>							
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
30	GM-CSF (myeloid)	-	-	+	-	5	GAS
<u>Growth hormone family</u>							
	GH	?	-	+	-	5	
	PRL	?	+/-	+	-	1,3,5	
35	EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)
<u>Receptor Tyrosine Kinases</u>							
	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
40	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., *Immunity* 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an Xhol site. The sequence of the 5' primer is:

5: GCGCCTCGAGATTCCCCGAAATCTAGATTCCCCGAAATGATTCCCCG
10 AAATGATTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

15 PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with Xhol/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTCCCCGAAATCTAGATTCCCCGAAATGATTCCCCGAAATG
20 ATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCC
CTAACTCCGCCATCCGCCCTAACTCCGCCAGTTCCGCCATTCTCCGC
CCCATGGCTGACTAATTTTATTTATGCAGAGGCCGAGGCCGCTCGGC
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTGAGGCTAGGCTT
TGCAAAAGCTT:3' (SEQ ID NO:5)

25 With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, 30 alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and Xhol, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter 35 element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using Sall and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 30,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to 35 generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells (10^7 per transfection), and resuspend in OPTI-MEM to a final concentration of 10^7 cells/ml. Then add 1ml of 1×10^7 cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Genticin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

10 On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

15 Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

20 After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

25 The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

30 As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced 10 in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest 2×10^6 U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM Na₂HPO₄.7H₂O, 1 mM MgCl₂, and 675 uM CaCl₂. Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then 20 resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

25 These cells are tested by harvesting 1×10^8 cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of 5×10^5 cells/ml. Plate 200 ul cells per well in the 96-well plate (or 1×10^5 cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma 30 can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes,

5 EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat phenochromocytoma cells) are known to proliferate and/or 10 differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

15 The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., *Oncogene* 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

20 Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes *Xba*I/*Hind*III, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

25 To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

30 PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

35 Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS

5 (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as 5×10^5

10 cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to 1×10^5 cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

Example 16: High-Throughput Screening Assay for T-cell Activity

NF- κ B (Nuclear Factor κ B) is a transcription factor activated by a wide variety 20 of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- κ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- κ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and 25 antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- κ B is retained in the cytoplasm with I- κ B (Inhibitor κ B). However, upon stimulation, I- κ B is phosphorylated and degraded, causing NF- κ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- κ B include IL-2, IL-6, GM-CSF, ICAM-1 and 30 class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- κ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- κ B would be useful in treating

diseases. For example, inhibitors of NF- κ B could be used to treat those diseases related to the acute or chronic activation of NF- κ B, such as rheumatoid arthritis.

To construct a vector containing the NF- κ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- κ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:
5' GCGGCCTCGAGGGGACTTCCCGGGGACTTCCGGGGACTTCCGGGAC
TTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:
10 5' GCGGCAAGCTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)
15 Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5' CTCGAGGGGACTTCCCGGGACTTCCGGGGACTTCCGGGACTTCC
ATCTGCCATCTCAATTAGTCAGCAACCATACTCCGCCCTAACTCCGCCA
20 TCCCGCCCTAACTCCGCCAGTCCGCCATTCTCCGCCCTAGGCTGACT
AATTTTTTATTTATGCAGAGGCCGAGGCCCTCGGCCTTGAGCTATT
CAGAAGTAGTGAGGAGGCTTTTGAGGCCTAGGCTTTGCAAAAGCTT:
3' (SEQ ID NO:10)

25 Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF- κ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- κ B/SV40/SEAP cassette is removed from the above NF- κ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- κ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- κ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the 10 following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15 μ l of 2.5x dilution buffer into Optiplates containing 35 μ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven 15 heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50 μ l Assay Buffer and incubate at room 20 temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50 μ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

25

Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6

23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO₂ incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO₂ incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10⁶ cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10⁶ cells/ml, and dispensed into a microplate, 100

10 10 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

15 To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca⁺⁺ concentration.

20

Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

30 Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members 35 of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

5 Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodynne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr 10 with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford,MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of 15 alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford,MA) are used to cover the Loprodynne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

20 To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodynne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na₃VO₄, 2 mM Na₄P₂O₇ 25 and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum 30 manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

35 Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

5 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg₂₊ (5mM ATP/50mM MgCl₂), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl₂, 5 mM MnCl₂, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the

10 components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

15 Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as

20 above.

25 Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity

As a potential alternative and/or compliment to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyn filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

25

Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., *Science* 252:706 (1991).

35

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and

5 Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated

10 according to Example 2 are nick-translated with digoxigeninideoxy-uridine 5'-triphosphate (Boehringer Manheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

15 Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv.

20 et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovation Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated

25 disease.

Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

A polypeptide of the present invention can be detected in a biological sample,

30 and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a

35 sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

5 The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

10 Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

15 Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale).
15 Interpolate the concentration of the polypeptide in the sample using the standard curve.

Example 23: Formulating a Polypeptide

20 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

25 As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 μ g/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If 30 given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μ g/kg/hour to about 50 μ g/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending 35 on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

5 The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

10 Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008;

15 20 U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

25 For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are 30 known to be deleterious to polypeptides.

35 Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as 5 ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, 10 manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

15 The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

20 Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

25 Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

30 The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 24: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

5 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily 10 dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

Example 25: Method of Treating Increased Levels of the Polypeptide

15 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a 20 polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

Example 26: Method of Treatment Using Gene Therapy

25 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is 30 turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

5 pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

10 The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to 15 transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

20 The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

25 Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is 30 required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 27: Method of Treatment Using Gene Therapy - In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense 5 DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide of the present invention. A polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the encoded polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, 10 see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) *Cardiovasc. Res.* 35(3):470-479, Chao J et al. (1997) *Pharmacol. Res.* 35(6):517-522, Wolff J.A. (1997) *Neuromuscul. Disord.* 7(5):314-318, Schwartz B. et al. (1996) *Gene Ther.* 3(5):405-411, Tsurumi Y. et al. (1996) *Circulation* 94(12):3281-3290 15 (incorporated herein by reference).

The polynucleotide constructs of the present invention may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). These polynucleotide constructs can be delivered in a 20 pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the 25 polynucleotides may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) *Ann. NY Acad. Sci.* 772:126-139 and Abdallah B. et al. (1995) *Biol. Cell* 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs of the present invention used in 30 the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major 35 advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct of the present invention can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, 5 uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is 10 similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, 15 although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of 20 DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of 25 nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for 30 delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding 35 for the polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with

liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made

5 on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel

10 clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from

15 different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice. The results of the above experimentation in mice can be used to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA of the present invention.

20 It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

25 The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

Sequence Listing

(1) GENERAL INFORMATION:

5

(i) APPLICANT: Rosen et al.

(ii) TITLE OF INVENTION: 86 Human Secreted Proteins

10

(iii) NUMBER OF SEQUENCES: 318

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.

15

(B) STREET: 9410 Key West Avenue

(C) CITY: Rockville

20

(D) STATE: Maryland

(E) COUNTRY: USA

(F) ZIP: 20850

25

(v) COMPUTER READABLE FORM:

30

(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

(B) COMPUTER: HP Vectra 486/33

35

(C) OPERATING SYSTEM: MSDOS version 6.2

(D) SOFTWARE: ASCII Text

40

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

45

(B) FILING DATE: June 11, 1998

(C) CLASSIFICATION:

50

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

55

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

5 (A) NAME: A. Anders Brookes

(B) REGISTRATION NUMBER: 36,373

(C) REFERENCE/DOCKET NUMBER: PZ008PCT

10

(vi) TELECOMMUNICATION INFORMATION:

15 (A) TELEPHONE: (301) 309-8504

(B) TELEFAX: (301) 309-8439

20

(2) INFORMATION FOR SEQ ID NO: 1:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 733 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA	GCCCAAATCT	TCTGACAAAA	CTCACACATG	CCCACCGTGC	CCAGCACCTG	60	
AATTCGAGGG	TGCACCGTCA	GTCCTTCCTCT	TCCCCCCAAA	ACCCAAGGAC	ACCCCTCATGA	120	
35	TCTCCCGGAC	TCCTGAGGTC	ACATGCGTGG	TGGTGGACGT	AAGCCACGAA	GACCTGAGG	
TCAAGTTCAA	CTGGTACGTG	GACGGCGTGG	AGGTGCATAA	TGCCAAGACA	AAGCCGGGG	180	
40	AGGAGCAGTA	CAACAGCAGC	TACCGTGTGG	TCAGGGTCCCT	CACCGTCCCTG	CACCAGGACT	240
GGCTGAATGG	CAAGGAGTAC	AAGTGCAGG	TCTCCAACAA	AGCCCTCCCA	ACCCCTATCG	300	
45	AGAAAACCAT	CTCCAAAGCC	AAAGGGCAGC	CCCGAGAAC	ACAGGTGTAC	ACCCCTGCC	360
CATCCCGGGA	TGAGCTGACC	AAGAACCAAGG	TCAGCCTGAC	CTGCTGGTC	AAAGGCTTCT	420	
ATCCAAGCGA	CATCGCCGTG	GACTGGGAGA	GCAATGGGCA	GCCGGAGAAC	AACTACAAGA	480	
50	CCACCCCTCC	CGTGCTGGAC	TCCGACGGCT	CCCTCTTCCT	CTACAGCAAG	CTCACCGTGG	540
ACAAGAGCAG	GTGGCAGCAG	GGGAACGTCT	TCTCATGCTC	CGTGATGCAT	GAGGCTCTGC	600	
55	ACAACCACTA	CACGCAGAAG	AGCCTCTCCC	TGTCTCCGGG	TAATGAGTG	CGACGGCCGC	660
	GACTCTAGAG	GAT				720	
						733	

(2) INFORMATION FOR SEQ ID NO: 2:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser
1 5

15

(2) INFORMATION FOR SEQ ID NO: 3:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 86 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCGCCTCGAG ATTTCCCCGA AATCTAGATT TCCCCGAAAT GATTTCGGG AAATGATTTC 60
30 CCCGAAATAT CTGCCATCTC AATTAG 86

35 (2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
45 GCGGCAAGCT TTTGCAAAG CCTAGGC 27

50 (2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 271 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
60 CTCGAGATTT CCCCCAAATC TAGATTTCCC CGAAATGATT TCCCCGAAAT GATTTCGGG 60

AAATATCTGC CATCTCAATT AGTCAGCAAC CATACTCCCC CCCCTAACTC CGCCCATCCCC 120
5 GCCCCCTAACT CGGCCCCAGTT CGGCCCCATTC TCCGGCCCCAT GGCTGACTAA TTTTTTTTAT 180
TTATGCAGAG GCCGAGGCCG CCTCGGGCTC TGAGCTATTG CAGAAAGTAGT GAGGAGGCCTT 240
TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

10

(2) INFORMATION FOR SEQ ID NO: 6:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32
25

(2) INFORMATION FOR SEQ ID NO: 7:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

40 GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

(2) INFORMATION FOR SEQ ID NO: 8:

45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

55 GGGGACTTTTC CC 12

(2) INFORMATION FOR SEQ ID NO: 9:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

10	GGGGCCTCGA GGGGACTTTC CGGGGACTT TCCGGGACT TTCCGGACT TTCCATCCTG	60
	CCATCTCAAT TAG	73

15 (2) INFORMATION FOR SEQ ID NO: 10:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 256 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

25	CTCGAGGGGA CTTTCCCCGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT	60
	CAATTAGTCA GCAACCATAG TCCCGCCCCCT AACTCCGCCA ATCCCGCCCC TAACTCCGCC	120
30	CAGTCCGCC CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA	180
	GGCCGCCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTG GAGGCCCTAGG	240
35	CTTTGCAAA AAGCTT	256

40 (2) INFORMATION FOR SEQ ID NO: 11:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1220 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

50	CATGAATGGC TCGCACAAAGG ACCCCCTCCT CCCCTTTCCCT CCTTCTGGGA GAACTCCCTC	60
	CCTCCCTCCA GCTCCGCCAG CCCAGGCCGC CCTTCCCTGG AAGCCGAGCG CCTTCGCTCG	120
	CATTTCACCG CGCCCGCCTC TCGCAATATT GCAATATAGG GGAAAAGCAG ACCATGGTGA	180
55	ATCCGGGCAG CAGCTCGCAG CCCCCCCCGG TGACGGCCGG CTCCCTCTCC TGGAAAGCGGT	240
	GCGCAGGCTG CGGGGGCAAG ATTGGGGACC GCTTTCTGCT CTATGCCATG GACAGCTATT	300
60	GGCACAGCCG GTGCCCTCAAG TGCTCCTGCT GCCAAGGGCA NTGGGGGACA TCGGCACGTC	360

	CTGTTACACC AAAAGTGGCA TGATCCTTTG CAGAAATGAC TACATTAGGT TATTTGGAAA	420
	TAGCGGTGCT TGCAGCGCTT GCGGACAGTC GATTCCTGCG AGTGAACCTCG TCATGAGGGC	480
5	GCAAGGCAAT GTGTATCATC TTAAGTGTTC TACATGCTCT ACCTGCCGGA ATCGCCCTGGT	540
	CCCGGGAGAT CGGTTTCACT ACATCAATGG CAGTTTATTT TGTGAACATG ATAGACCTAC	600
10	AGCTCTCATC AATGGCCATT TGAATTCACT TCARAGCAAT CCACTACTGC CAGACCAGAA	660
	GGTCTGCTAA AAGGTCAAGAG TAATGCAGAA TGGCTGCCTT CATCTCAGAT TTGTTCATCA	720
	CAGGTGGATC CCAATGTTCTC TCAGTAGACA AGTCACCTTT GTAGCTAGCA CCAGTGCCAG	780
15	CTCCATGCCA TTGACACCTTC TTTAGTCCTTG ATTGCCCTTC CCGCATTIWT TGGTGTATTA	840
	AAATGACTRA TKAAGCTAAT TAAAAGAAGC ATTCAAATCT GCTTTCTACC CTCATTAACA	900
20	ATTAGCAGGG CACTGGCCAG AGTTTGTACC CTGTGTTTTA CCTTAACAAAC ATTCTATTTG	960
	CTCTTGTAT ATTAAAGTGT TGTAAGGAAA CGTGTTCAA TCAAAACTGA CCATGAGATA	1020
	AAGGAAAGAG ATGTGGCTTT TGTGATATTTC TATCACAAAC ACTTTATTGTA TCTCTGTAAA	1080
25	ATACAATGTA TGTATGCATG TAAGTGTTC TGTCTTAATG TTGCTACTCC CATGGCAAAG	1140
	AAAAAAAAAA GAATGAAAAA ARAAAAAAA AAAAAAAA AAAAAAAA CTCGAGGGGG	1200
	CCCCCGTACCAATGCCCT	1220
30		

	(2) INFORMATION FOR SEQ ID NO: 12:	
35	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1939 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
40	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:	
	GAACACAAAC ATGCAGTCTG TAGCAGATGG TAATAGGCTG AYATATTACA CTTGTTGATG	60
45	TAAAATCTGAT AGGTTTCATT CTCTCCAAGG ACAGCTTTTT AAATATTAA CAGTATCAAT	120
	AATTTCAG TTTCTGTGAG AATTTATAA TTTATAATTG GCAGACTTAA TGTATAATCT	180
50	ATTTTGTCTT AACAATTACA AATATATTTT TTATTTCAGA TTATATATAT TCCTACCAGA	240
	TGGAGATAAT TACAGCTTTA AAAATTTTA TTTTTTCATT TTATTTCACA CATTGACATT	300
55	AAATTTTTAT GGACACATAA TAACTGTACA TATATATGGG GTAGAATGTG ATGTTTTAAT	360
	ACATGTACTC AATGTGTAAT GATCAAATCA GGGTAATTG CATAATGATT TTTCTGTAGG	420
	GAGAAAATTC AAAATCTACT CTTCTGGCTA TTTCAAATA TATAATATGT TATTGTTAAC	480
60	TATACTCATC CTACTATGCA ATAGGACACC AGAACATTATT CCTGGGTCT ACATCCGTTA	540

	AGGCAACCAA GGATTGGAAA TATGGAAAA AAAAATTGCG TCTGTACTGA ACATGTACAG	600
5	ACTTTTTCT TGTCCCTTATT CCTTACACAA TATAGTACAA TAACTATTG CATGACATT	660
	ACATCGGATA TTATGAGTGA TCTAGAGTTG ATATGAAGTA TATGGGAGGA TGTGCAAAGG	720
	TGATGTGCAA ATACTATGTC ATTTTATATC AGGGACTTGA GTATCCTTGT TTAYCCTCAG	780
10	GAGATCCTGA AACYAGTCCC CCATGGATAC TGAGGGCTGA CTGTATAGTC CTATCCTCAC	840
	CGAACCTTCA TTCTAATGRG GGAAGACTGA CTATAAACAA AATATATGTA ATAGGTGGTG	900
15	GTAAGTACCG TGGAGAAGTA ACAAAATGGGG CAAAGTGAGT TATACAGTC CATYCTAGA	960
	AACCTTGGAG TACTTTCTT AGTTTATACT CGTGGTGGTT TCCCTTTGTC TCCCTTATT	1020
	CATGGGACTC TGACATGTGC CCATAGCTAG GGTGGCAGTA GGATCTACCC GAAAAGCGTC	1080
20	CTGCTGATAC AGGACCAAAG CATCCTGTTG TTCTCGAGCC TATAAAAAGA GCTAATGGTC	1140
	TTGCTTCTCT TAACTGTGGC CTCCCTACACT GTGTTTGGGA TGATTGGTGA TGTCTTGGAT	1200
	ATTCCTGTTT TTTGGAACCTT TGAATATACA ACACTTTACT AGGAAATTAG CAATGGAAGC	1260
25	AGAGCAAAGA TGTACAGAGG AAACAATGCR TAACTCTGAT GGAATTGAAG TCATGAGGCA	1320
	GCAGAGAGCT TAAATTASAG CTTTAAAAAT TTTTATTTT TAGAGGGAAT TTAMTTGGGA	1380
30	GTAACAGCAG TAATAGTTAA CGGAGCCAGA ATGCTTGAGT CATATAATTG CAAAGCAGAG	1440
	TTGGGAGCAA CAGATGCTAA AGAGTAGTTG CTGTAGTTCC TCTTGGGTC GTAGGAGCAG	1500
35	TTGTCATTTT MCTATAYAGC TACTGCATGA AGAAGAGTTC TTAGTGAGGC CTGGGTGAAC	1560
	AGCTCTCTT AGTATTCTGT GTGACCCCAT TYGACCTTTT ACAAAATCCC TAAGTAAATA	1620
	AATAGCCCCCT MAGGWAACACT AAGTTTTCT CTGCTGTTT TTTGCTTGAG AGAGCTATAA	1680
40	CTGTAATAGA CTTATATTTG TGAACATTTC AGTGCTTGCC AATATTTGGT AATATTTATG	1740
	TTTCCTATAT TTGTAATGAA CATTCTCTT CMGGTACATT TTGTTGAAA TTATGTTT	1800
45	ATGSATAAAA GTTCACCTTT TATTGTATAA AATTGACTCA GATTAATTAA TACACATTGA	1860
	CAATGGGTAA ATAGAGTTTT TCAGATTATT AAAAGCTGAA GGATGCCCAT GAAAGCAAAA	1920
	AAAAAAAAAAA AAAACTCGA	1939
50		

(2) INFORMATION FOR SEQ ID NO: 13:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2602 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

	GGGTCTTCTGG CGCAACTTTC CTTTCGGGT GTTCTGAAGC GGTTTCTCTG TAATCCTCAG	60
5	TGAGGAAACC CACCGTGAAT CGGATTGCCG TTCAGTCCC CGGAAGCCTG GCTCGTTGGC	120
	CATGTNGGGG ACCCATGTC ATTAAAGTCA TTAAAATAAT TTCAATTGTC TTGGTTGAA	180
10	GACTGCTTCA TTCTGCCCT AGTACCAGCG GTTCTCTGT TCTGTGATCA ATGTGATTCA	240
	CAGGAACCTCC TTAAGTAACA AACGAAATGA GCCAGGGCG TGAAAATAT GACTTCTATA	300
	TTGGTCTGGG ATTGGCTATG AGCTCCAGCA TTTTCATTGG AGGAAGTTTC ATTTTGAAAA	360
15	AAAAGGGCCT CCTTCGACTT GCCAGGAAAG GCTCTATGAG ACCAGGTCAA GGTCGCCATG	420
	CATATCTTAA GGAATGGTG TGGTGGCTG GACTGCTGTC AATGGGAGCT GGTGAGGTGG	480
20	CCAACCTCGC TGGTATGCG TTTCACCAAG CCACCTCTAGT GACTCCACTA GGAGCTCTCA	540
	GGGTGCTAGT AAGTGCATT CTTCTTCAT ACTTTCTCAA TGAAAGACTT AATCTTCATG	600
	GGAAAATTGG GTGTTGCTA AGTATTCTAG GATCTACAGT TATGGTCATT CATGCTCCAA	660
25	AGGAAGAGGA GATTGAGACT TTAAATGAAA TGTCTCACAA GCTAGGTGAT CCAGGTTTG	720
	TGGTCTTGC AACCCCTGTG GTCATTGTGG CCTTGATATT AATCTTCGTG GTGGGTCCCTC	780
30	GCCATGGACA GACAAACATT CTCTGTGACA TAACAATCTG CTCTGTAAATC GGCGCGTTT	840
	CAGTCTCCTG TGTGAAGGGC CTGGGCATTG CTATCAAGGA GCTGTTGCA GGGAAAGCCTG	900
	TGCTGCGGCA TCCCTGGCT TGGATTCTGC TGCTGAGCCT CATCGTCTGT GTGAGCACAC	960
35	AGATTAATTA CCTAAATAGG GCCCTGGATA TATTCAACAC TTCCATTGTC ACTCCAATAT	1020
	ATTATGTTATT CTTTACAACA TCAGTTTAA CTTGTTCAAG TATTCTTTT AAGGAGTGGC	1080
40	AAGATAATGCC TGGTGACGAT GTCATTGGTA CTTTGAGTGG CTTCTTTACA ATCATTGTGG	1140
	GGATAATTCTT GTGCAATGCC TTTAAAGACG TCAGCTTCAAG TCTAGCAAGT CTGCTGTGT	1200
	CTTTCTGAAA AGACGAGAAA GCAATGAATG GCAATCTCTC TAATATGTAT GAAGTCTTA	1260
45	ATAATAATGA AGAAAGCTTA ACCTGTGGAA TCGAACACA CACTGGTGAA AATGTCTCCC	1320
	GAAGAAATGG AAATCTGACA GCTTTTAAG AAAGGTGTA TTAAAGGTTA ATCTGTGATT	1380
	GTTATGAAAGT GAATTGAAAT ATCATCAGAA TGTGTCTGAA AAAACATTGT CCTCAAATAA	1440
50	TGTTCTTAA AGCAATCTT TTTAAAGATT TCACTAATTG GGACCAAGAA ATTACTTTTC	1500
	TTGTATTAA ACAAACAAATG GTAGCTCACT AAAATGACCT CAGCACATGA CGATTCTAT	1560
55	TAACATTAA TTGTGTAGA AGTATTTCAC ATTTCATCC CTTCTCCAAA AGCCGAATGC	1620
	ACTAAATGACA GTTTAAAGTC TATGAAAATG CTTTATTTCAG TCAATTGGTGA TGAAAGCTG	1680
60	AAATGTCAT TTGTCAATCCC CACTCCATCA ATCCCTGACCC ATGTAAGGCT TTTTTATTTC	1740

	AAAAAAACAG AGTTATCCCA ATACATTATC CTGTGATTAA CCTTACCTAC AAAAGTGGCT	1800
	CCTGTTGTT TGATGATGAT TGGTTTATT TTTGAAATAT TTATTAAGGG AAAACTAAGT	1860
5	TACTGAATGA AGGAACCTCT TTCTTACAAA ACAAAAAAAA GGGCAGAAAT CACCCCAAGG	1920
	AACGATTCTC CAGGTTGAGA TGATCACCGT GAATCCGGCT TCCTCTGAGC ATTCGATGGC	1980
10	CTTAGCACCT CATCAAGCCA GCACATCCTG CCTGCTGTTG CAGCCTGGCT GGGTTTATTC	2040
	TTCAGTTACC CTAATCCCAT GATGCCCTGGA ACCTTGATTA CCGTTTACA TCAGCTCTTG	2100
	TACTTTTCAG TATATTTCA TAATGAGTTA TATTGTCATT TAGACTTTGA ACAGCTCTGG	2160
15	GAAATAGAAG ACTAGGGTTG TTCTTAAAT TTAGCTCATG TTATAATAAA AAGTTGAAAT	2220
	GAAGTTCTTA TTCTAAAAGT CTGAATGCTT AGAACAAACT TAACATGTTT ATAGAATATG	2280
20	GTCTCTTGT ACCAAGTACT TTGCTTAAGA GCTCCTTGG GCCACTACAT ATTTGGTTT	2340
	CTAGAAAATG TTGTTTATG AAGAAGTCGA TGGAAAATC CAAACATATG CAGAAAAGGT	2400
	AGAATAATAA AAAAGGTCTA ATGAACTCCA TTCAGCTTGT AACCTATCCA CTCATAACCA	2460
25	TTGACTGGCC TTTTAAAAAA AAGTATTGGG CAGAATTAAA TTTCACCTA GGTGATGGGG	2520
	AAGGAAAGTG TTCCCTGTN CCAGCCTGTG GTTCCTGCCT GGGNGTTTA CCCAGTGGTG	2580
30	GGCCCAAGGCC AAGGTCCATT CA	2602

35 (2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 808 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

45	ACCCACGGGT CCGGTTAAC AAAGGGAATG ACAGATATGGG AAAGAAAATA CATTGGATG	60
	TTACAGATAT GTGTGTTCTT GGAGCCCAAGG GCCAAGCCCT CCCTGGGGGA CTGGATGG	120
	TGATCTCTCT CCTTGGCCCC AACCTGACAT CTTTCTTGT CCTTTAGGA ATGTCTGATG	180
50	GAAATTCCCTC CTAACCTGGG GTCACTACTCC ATTCATTCCT CTGGCTCAN TGAGAAGGAA	240
	AATTTTTTTT TAAGTAATT ACTGAAAACC CAGATCACAC CATATAAAAT TCAGATAGGT	300
	GCAATTCTGC CCACAATGAA GGCAAAAGTGT TACACTAATT TGAAAACAGT TTAGCCTCTT	360
55	AATCCCCCAA ACTTCATTCT TGAATTTGT CATTTCCTGT GGGCAAGCTG TGGGAAAGGG	420
	GCACAAAAGT ATCACTGAAG TATTTTTCA AAAAGAAAA AAGGCAGTCT TCCCTACTA	480
60	ATGAGAATGC AAAATGTTGA ACAACTGTAA AATGTTTCA CCCTGCTTT AGACATAAAG	540

	CTTTAAAAAA CTGTGAGGTC TTTTATCACT TCCCCATTGT ATATGTAATA TGGCTCCAGA	600
5	TAATTACTCT GCCACGGGGA GAAAATCTTC CATAACTCTC CCCTATATAT ATGTATACTC	660
	CACCACTTA TCTTGTATG TCATGGTGGT GGGAGTATTT ATMCACAGA AACAGGAAA	720
	TGATACAAAC CTGGCGACA GAGCAAGACT CCACTTCAA AAAAAAAA AAAAAAAA	780
10	AAAAAAAAA AAAAAAAA GGGCGGCC	808

15 (2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 864 base pairs
- (B) TYPE: nucleic acid
- 20 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

25	GGGTTTTTIG TTTTGTGTTT TTNAGGGGG AGGGGGGGT TCCCCCTCCTT TGCCCCAGAC	60
	TTCTCTTGA ACACAAATGC ATTAGCCTTG TGGCTAGAAM ACCCTCTTCC TACCTCTGTC	120
30	TCCCCCTCACT TGTCATATGC TCTGACATGC TAACATTTCT TTTGTTCACT CCTGTTGCC	180
	CCACAGAAAC ATCCCAGAAA AACCGGTCAAG TGTTCCTTCC TCCCTGATCC TTAGGTTCT	240
	GAAATAGGGT TCTGTTACAT CCTCTTGGAT AGCCTGTTA AAATGTTAG AAGGCTGG	300
35	GCTCAAAAT GCGTTCTTCC ACATTGATAA TTAGTAAAC TGAGAACATT GACATCACTA	360
	CAGGGCAGCA TAAGAGGTTG CTTACATGTG GTAGCAGCTC TGGTTTGATT CAAGTTGCTA	420
40	CCATGTACAT TGACAGCACA TATACCATAA CCAGCGTGTG GGGTTGAATT GCACTTTCTA	480
	CCTTGTATG AGATTTACAG ACTTTCTTC TGGGTTGTA TCATGACCAAG AGGGGTACTA	540
	TAGGGTTGGT TTATACTGCA ATATAGAGGA TCAGAAGCCA TTTGATTGG TAGGTGTGTC	600
45	AGAAGGGAGA ATGATGGCAG ACGAACTGCT GGAAGAGGTC AGAAGATAGC CATGCTAAA	660
	TGCAATTATA TCCTCATGTT TATCCAAAC TAATCTTGGA CTTTTCCACT CATTAGCTTT	720
50	GTGTTGCCCCCT TGGTTCCCTT GAAGGTTAA GTTCAACCAT ATTCTGICAA CTGTTCACT	780
	TCAGTGGAAAT CTTGTATTTC TGGTTCATTA TAACAAATTG TTGCGTTAAA AAAAAAAA	840
	AAAAAGGGCG GCCGCTCTAG AGGG	864

55

(2) INFORMATION FOR SEQ ID NO: 16:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2361 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GGCACGAGCT CGAGTTTTTT TTTTTTTTTT TTCTTATTTT TGCCAGACTC TTGATACTCT	60
10 TAAAACCTGT TTGTTGGTCAG CACAACAAGG AACAAAACAA AGCTTTGAAA AAACCTTTAAC	120
ATGAAAAAAAC GCACGTGACAT TTTTTTTTAT TTAATATAGC CTGGACTTTA CCTGCGTATG	180
15 CACATGCTCA GAAATTGTCTA CTAGGCTGAC TATGTATCAC CTCTTCAGCT TGGATCCAAT	240
TGTTGGATTTA TTACAAACA TCAAATGCCCT TCAAGCCAACT CCTTTTTGCT GTATGTTTG	300
CACCTTACTG TAGTAGATAC GCAACAGATA WTGTGGAAA AAAAGAGATA AGAGGAGGAA	360
20 GCTAATAAAGA GACTGTCAAG ATTGTATACC TTCTTGGTTT CTTTTAAAGAA TTGTGTTGCC	420
TTCTACTATT ACAGCAAAGC AGCATTGTGT TACTGACTGC CTAAAATCAC TTAATCTCAG	480
25 GTGAACGCAT CACTTGCCTA ACTGTTGGAA TGCTATTGTG TTGTTGTTGC ACTGTTTTTT	540
TCGTTTGTGTTT GTTTGTGTTAT TTGGTTGGCT TTTTGGAGAG GGAAATTGAA AACGGGACA	600
TACACAAAAG TTACACACCC ACATTCCCTT TTTATCATGA CATAACAAGAA GAAACTAGCA	660
30 GACCTAAGAA TGGAGTGAAG AAAGGCAGTA TGGCAGGCAC CAGCAAAGAG TTGAGGGCTG	720
TTGCTCTTAA AAATTATTTT TTTTATTATT ATTGTGAAAG TATGGAAGTT TTCCATTAC	780
35 TGGGAAAGG AGGGAAAAGT GCATTTATTT TTATACAGAG TTACTTAATT ACCTCCAAA	840
CACATATGTT GGAAATCGCT TTGCTGGTG CAAAGTATAT TAATGAGCAG GAATACATAC	900
ATGGAGGTTA TGAATAGAGA GCTCAATTG TACCTTTGCT GTCTTGCTCA AGCTTGGTAT	960
40 GGCATGAAA CTCGACTTTA TTCCAAAAGT AACTTCAAA TTTAAATAC TAGAACGTTT	1020
GCTGCGATAA ATCTTTGGA TTTTGTGTT TTTCTAATGA GAATACTGTT TTTCATTAC	1080
45 TAAAGAACAA TTGCTAAAC ATGAGAAATC ACTCACTTTG ATTATGTATA GATTACATAG	1140
GAAGAACAAAT CACATCAGTA AGTTATAGTT TATATTAAG GTAATTTCCT GTTGGCTCAT	1200
AACAAATATA CCAGCATTCA TGATAGCATT TCAGCATTTC CCAAGGTACC AAGTGTACTT	1260
50 ATTTGTGTTGTT TGTGTTGTT GTTGTATTTT AGAAGGAATT CAGCTCTGAT GTTTTAAAG	1320
AAAACCAGCA TCTCTGATGT TGCAACATAC GTGTAAAATG GGTGTTACAT CTATCCTGCC	1380
55 ATTTAACCCC ACAGTTAATA AAGTGGCTGA AAATAATAGT AGCTCTGGCT TGGTGGCTGA	1440
CCTGGTTAAA TACTGCTTAA AAGCTCATAAC AAAACAAATA GGCTTTCCA TAAGTGGCCT	1500
TTAAGAAAAC ATGGAAGACA ATTCACTGTTT GACAAATGCT GACAGGGTGA AGAAACCCCA	1560
60 GTGTAAAAAT GAATCGCGTT TTAAGTGATT CGGTTAAAGA GTTTGGGCTC CGGTACCAAA	1620

	CTAATACTAG ATAATAAGGA AATGGGGTG AAATATTTTT TTATTTGTGA ATCATTTGT	1680
5	GAATGTCCCC CTCAAAAAAA CTTAAATGGAA TATTTGGCAT AAAGGGCATT TGGTGGTTTT	1740
	ATTTTTGTAA GAGGGGGWTT GTCAGAAAAT CCCTTTCTC TCCTTACGYCT AACTGACTAG	1800
	GGAAACAATTG TTGATATGCA TAGCATGGG AACTACTGTC ATTATATACT CTTACAAATA	1860
10	ACACATGAAG CAAGAATGAC CAATATTCTG NATAATTGGG CACTGGGATC ACAAAATGTG	1920
	ATAAAACMTT AAAATGTATAA AACTTTATCA AATAAAAGTTT TATTTTCCCC TTTAAAATGT	1980
15	ATTTCTTTAG AGGCATTACT TTTTTAAAAA TATTGGTCAA TTCTGACAT AAGATGTGAG	2040
	GTTCACAGTT GTATTCCAGT ATTCAAGATA GATTCCCTGAT TTTTCAATTAA GGAAAAGTAA	2100
	AATCCAAAAT GTTACCAAAA CAAAGTCCAA TATTAATGT TTGCTTTATA GATTATATTC	2160
20	TATGGCTGTT TGTAAATTCT CTTTTTTTCC TTTTTTATTT GGTGCTGAAT ATGTCCTTGT	2220
	AGGCTCTGTT TTAAGAAAAC AATATGTGGG AAATGATTAA ATTTTTCTA TTGCTCTTCC	2280
25	TTGTGGAAAA TAAAGTGTAA TTGTTTTTTC TGTTTTGTAA AAAAAAAA AAAAAAAA	2340
	AAAAAAA AAGAANGAGA A	2361

30

(2) INFORMATION FOR SEQ ID NO: 17:

	(i) SEQUENCE CHARACTERISTICS:	
35	(A) LENGTH: 803 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:	
	CAGCTGCCCA CAAGGTGGGC TCCTGGGGGA GGGTCATCCC TCTGAGAAGA GGGCGGCACC	60
	AAGACCCACA CACCTGAAAA ATGTGGTACT TCATGTCGCT GATCTCGATG GTCTTGCTGC	120
45	TGTCCCCATC CTGTTCTGAT TTATTTGTC TTAGTGTCTT GAACCTGGAG CAAAGGAGAC	180
	AAAGCAAGGT GGGTTTGAA CCTTTTACTT CACCACTGTG TGGCGNATGG CACCATCTGT	240
50	CACCTGACCG GCTACCACAA GACGGAACAT TTAAATT ACTGCTGTGC TCCTAAAATA	300
	ATTTTCAGCA AGTGCCTATT TACACCATCT TAGGAAGACA TCTGAGCTGA GCCCAATTCT	360
	GTCCCCACCA CCCACCCCTAC AAGCGACCTG ACGCCTGTGG CCAGAATGCT GACTCTTCAT	420
55	TCCAGGATAT TTATGTTTTC TAATAATAAA AGCAATAACT AGGCCAGAAA GAACACCACC	480
	TCAGAGCCCC CCTTTCTGTC TGCCCTGGGT CCACCCCGTC TCATCCCCCT GTGGGGCGAG	540
60	TGGGGCTCTG CTGCAATGTG ACTGCAGTCT GAGGGCCAGA RGCTGCAGGK TACAGCCCCA	600

	GGCARTCACT CTCTGTCACC TGGAATCTGA AACAAAGGTGC TTCTGTGCCCTCTGGG	660
	AGTTTGTAT CTGAGGCTGC CTACCTGTTA GAACNTGTCA CCAGCAGGAC TTTATGTGCA	720
5	TAAAACAGCT TTCCCTCCAC CAAAAAAA AAAAAAAAC TCGAGGGGGG GCCCCGTACC	780
	CAATTCGCCCT TATAGTGAGC GAT	803
10		
	(2) INFORMATION FOR SEQ ID NO: 18:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 1794 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:	
	TTCCTTTTTCG TTCATGGGAC ATGGTACCTA AGCAAATAGG AGTTGGGTTT GGTTTTTCTC	60
	CTAAAATAAT GCTCAAACT TACCTAATCA AATGGCATCC ATTGAAATAA AATGACAATA	120
25	ACTAAAGCTA GTTAATGTCA GTGACATTAA ACTAACTCCA GGATTTCAGGA GTTTTAATGT	180
	TAGAATTTCAG ATTTAACAGA TAGAGTGTGG CTTCATTTGT CCATGGTAGC CCATCTCTCC	240
30	TAAGACCTTT TCTAGTCTGT CTTCCCTGCCT TCGAACCTGA TGACAGTAAA ACCCTGTTA	300
	GTATTCTCTT GIGCATTGG TTTGTTGGTT AGCCGACTGT CTTGAAACTA TTCATTTGC	360
35	TTCTAGTTT ATTTAACAGA GGTAGCAATTG GTGGGTTTTT TTTTTTTTTT CTGCTCTGT	420
	GTTTGAAGTT TCAGTTCTG TTTCTAGGT AAGGCTTATT TTTGATTAGC AGTCAATGGC	480
	AAAGAAAAAG TAAATCAAAG ATGACTTCCTT TTCAAAATGT ATTGTTAGC ACTTAACCTCA	540
40	GATGAATTAA TAAATTATTA ATCTTGATAC TAAGGATTIG TTACTTTTTT GCATATTAGG	600
	TTAATTTCATA CCTTACATGT GAGAGTCCTA CCACTAAGCC ATTCTGCTC TGACTGTG	660
45	GGAAGTTTG GAAACCCCTG CCAGTGATCT GGTGATGATC TGATGATTTA TTAAAGAGC	720
	CGTTGATGCC TCCAGGAAAC TTAAGTATT TATTAATATA TATATAGGAA TTTTTTTTA	780
	TTTTGCTTTG TCTTCTCTC CCTTCTTTTA CCCTCATGTT CATTCTCAA ACCAGTGT	840
50	TGGAAGTATG CATGCAGGCC TATAAATGAA AAACACAATT CTTTATGTGT ATAGCAATGT	900
	TATTAATGTC TAACTACATA CGCAAAACT TCCCTTACAG AGGTTGGAC TAACATTTCA	960
55	CATGCACATT TCAAAACAAG ATGTGTCAATG AAAACAGCCC CTTTACCTGC CAAGACAAGC	1020
	AGGGCTATAT TTCAAGTGACA GCTGATATTG GTTTGAAAG TGAATCTCAT AATATATATA	1080
	TGTATTACAC ATTATTTATGA CTAGAAGTAT GTAAGAAATG ATCAGAACAA AAGAAAATT	1140
60	CTATTTTCAT GCAAATATT TTCACTCAGTC ATCACTCTCA AATATAAATT AAAATATAAC	1200

5	ACTCCTGAAT GCCTGAGGCA CGATCTGGAT TTTAAATGTG TGGTATTCAT TGAAAAGAAG	1260
	CTCTCCACCC ACTTGGTATT TCAAGAAAAT TTAAAACGAT CCCAAGGAAA GATGATTGT	1320
	ATGTTAAAGT GACTGCACAA GTAAAAGTCC AATGTTGTGT CCATGAAAAG GATTCCCTGG	1380
	TTATGTGCAG GGAATCATCT CACATGCTGT TTTTCCTATT TGGTTTGAGA AACAGGCTGA	1440
10	CACTATTCTC TTGATTAGA AAATAAACTC ATAAAACCTCA TAATGTTGAT ATAATCAAGA	1500
	TGTAACCACT ATAAATATGT AGAAGAGGAA GTTTTAAAAG ACCTTAAGCT GGCATTGTGA	1560
15	AGGAACACCA TGGTAGACTC TTTTTGTAAA TGTATTTGT ATTAAATGAA ATGCAGTATA	1620
	AAGGTGGTG AAGTGTAAATA TAATTGTGTA AACAAATCCT GTAAATAGAG AGATGTACAG	1680
	AATCGTTTG TACTGTATCT TGAAACTTGT GAAATAAAAGA TTCCACCTCT GGTTAAAAAA	1740
20	AAAAAAAAAA AAYCGGGGC CAGTTCCCCC CGGGCTATTT TAAAAGGNAAA AAAG	1794

25 (2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1037 base pairs
- (B) TYPE: nucleic acid
- 30 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

35	TCGAGTTTTT TTTTTTTTTT TGACAGAGTC TTGCTATGTG GCCCAGGCTG GAGTGCAGTG	60
	GCAATCTTGG CTCAYTCAA CCTYTGCYTC CTGGGTTCAA GCAATTYTCC TGCYTCAGCY	120
40	TCCYTAGTAG CTGGGACTAC AGGCACCTGC CACCATGCCA GGTTAACCTTT TTGTATTTTA	180
	GTAGAGACAG AGTTTCACCA TGTGCCCCAC GCTGGTGTGG AACTCCTGAG CTCAGGCAAT	240
	CTGCCACCT TGGCCTCCGA AAGTGCTAGG ATTACAGGCT TGAGCCACTG CACCCAGCCA	300
45	AGCTGTACTT TTTTTTTTTT TTTTAAAGCT TCAAACCTTC AATATTTCAT TAAGAGTTAC	360
	AGTTGGTTT CAGTCATTCK GAGGRAAATT AAGGAAGGGG CTTGGCCCAW ACCTGGTAAA	420
50	AGAATGGAAG GAACCAATT TTAACCATTG GGACCACTG AATTCATGG GAGTGCCTTT	480
	TGTCCCCCAG GAAACATCTR GAAAGGTATA WKAGATATT TSTGGSTMGT CACAATTGT	540
	GATGGGGAA AAAAGAACTA CCAGTATCAG GGGGATACAG GCCCCGGTATC AGGTGGATAG	600
55	AGGCCTGGAA TATTGCTAA CATTCTACAG TGCAAAGACA SCCTTTMACA WACAGAACTA	660
	TYTGGTCAA AATGTCATAA GTGCTGAGGT TGAAGAACTC AATATTTAT ATGTTTCAG	720
60	GGAATTCTA TGTGGGCTTG GGAAAGTTTG AAGTCATTG TCATTTGTAT ATTTAAAGGG	780

	ATATATTTTA TCATTAGTCT ATAAATTCCA GTGCCAAAGT AGAGGCCCTG CACATTGTG	840
	CACATATACA CACACCAGAA ATAAAYTMTTC TKGCAATTAT CTTCTCTATC ATTGACAGGG	900
5	CAATGACCTA TGAAAATTAT GTTATGTCTA ATAGTCCCTC ATTGTTATGT GCAAAACACC	960
	CAGCAAAGCT CAAGTTAAGR TTGTGGTCAC AAAGAAAAGA GCTATCATTG CTTTATGATG	1020
10	TTGCTCTGAAG TTAATGA	1037

15 (2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1309 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

25	GGCACAGACT TTAAGAAATG CCAAATGCAA GGACCATTAA GAAAATTCTC CCCGAAATGA	60
	GGCTCCTCTA ACAAAATGATG ATTANAACGC TCTCTCCCTG ACCAGTCACA TTCTAGAAAC	120
	ACGACATTCC ATGAGGCAGG AAGAGTTCAAG TTAATTGCT CCKGAAAAAG TGTGGTCAG	180
30	TGTTTGTGTG GCAATGTACG TGGGCAGAAG AGGCCGCTCA AGCTGTGTCC CCCCTGAGCA	240
	GGATTTCAGGA AAGGGAAAAG AAGTCTCTT CAACTCAGCC AAGGGGCCGT ACGATGGCCG	300
35	ATGAGATTAT GTATTTAAAA GTCTTTGTA AAGTGTAAAC TAAAAACCTT AAATGTAAAGA	360
	TGCTGTTGTT ATTATTACTG TTGTTGTTGC TGTATGGAC ATGCCAAAAG GCCCTTGTAA	420
	GAAGACAGTT TTGCTTTTC AATCTCATAG CAAGGAACTC AAGTCTGAAG CTTCAAAAG	480
40	ATGAGAAGAA GGGCAAGAAG AGGGATAACT CCCAAGCTCA GAGGGAAAAA AAAGGTGGGG	540
	GAAAAGAGCC CCAGGGTGAC CTTCAAGAAA GGCCAGGACCC AGGATGATCT AACCTTTCCC	600
45	TTCACCAGAA ACAAAAGCTAT TGCCAGACTG AACCTTAAAG TCAAGCAGTC ACCCACTGCC	660
	TTTGTGGGA GCAGAAGGCC ATAGCAACAA GTGACCTGCC CCTCAGACCTC AAGATCCAG	720
	ATACCAGAGC TGGAGGAGTC ATAGGGCATT ACTGGTAGGC AGGAAAACGT AGGGTCGAAC	780
50	AAATGGAAGA ATGCGGTGAT CATAGACCAA AGACACACAG ATAATTAACC CCATGTGTCC	840
	ACCCAGGCCA AAGTCTTCC TGCTACCCCA CAGTGGATGT CCAGGCAGAT GGTCCCCACA	900
55	TGATGGGAA GCAGAGGGCA TAGTGTGGTT TTGTGGACT TGTCATGTT TTGTAGTGTG	960
	GGCTAACAG TGCCAAAGGA AACACTAGGG AAAAGTTGGT GAAACATGCC AGCTAGCAGG	1020
	ACCAGTAAAG GCATAATCAG GCATTTGGCA AAGCTTGCTT TTCTAATTCA ATGATAGGTT	1080
60	CTAATAGGAA ATTTTTGAAG ATTTTTAAA ACAATGTAT AGTGGCAGTCC CCCAGTATG	1140

GAATAAATAA CATGCATTCT TTTTCAATA TACTGTCTA TTTCAGATGTC ATTAAAATAA	1200
ATGGATGAGT CACAGAGGAG CTATCAGATG CTCTCATGAC TACCTAACT CAAAAAAA	1260
5 AAAAAAAWA AAAGGGGGGC CGGTACCCAT TTGCCCTAAA GGGATCGTA	1309

10

(2) INFORMATION FOR SEQ ID NO: 21:

15 (i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 1081 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: double	
(D) TOPOLOGY: linear	
20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:	
ACANATTTT TACTTAAATT TTATTTTATC TTATTTTTAG GTGCTTTAA TCTAAAATT	60
CTGAAAAGCG AATAGCACGT GTTTTCAGAA ACAAAATGTGA AAGCAGTCAA ATTAAGTAGA	120
25 TACTATTTAG AAATGTAAAA TACTCTCCAG ATCTACCATTT AATAGAAAAT AAACTAAACC	180
TTATATTTTA TTTTGCCAA AATATTTTAT TATAAAATAT GACAAAATA TTTAAAATGC	240
30 ACAATGCTTT TAACITAAAT GTGCTAACCC TGTTCCTGTC TGTTCCTGTC TGTACCTTT	300
CTGATTGCGA ATTATAGAAA ACTTGATAAA TACTTGATTT TAACCAATGA GACTACAGGC	360
AGATGGGACT AAGTGTTTAT GGGACAATTA TGTACTATTT AACTTAAATA TATTTTGTT	420
35 AATAGGAAAT ATATAATAAT AGCATTATAT GTAATAAAAT ATGGGCAACG ATTATCTTGG	480
AAATTAAGA GTCAAAGCAA AGAAAATGAAG GGCTGGTAAA ATGAATTTCG TAATATCCTC	540
40 AGGATACTTT TACCTTAAAA GTATGTTGTT AAAGATTTTG TAAATTGTAT TTCAACAATT	600
TTAAATGTGT TGAGCAAGTT GCAGTGCAAA CACTGTCTT ATGTAGAGAG TTTATATGCA	660
CATAATAACC TGTACCTATA AATCGTGCAA TAACCATATG CGACTATTTT GCCATGGAGA	720
45 AATCTGACAG CATTGCAAAC AATAGTATTG TTTGATGTAG TTAACCTAA GTTATTTTC	780
AGTAATTCTC TCACAAATCA AGATTCAAAC AGCTTAAAC ACTTCCAATG AGATAAAATA	840
50 TTTACTATTA TGCTTATTAG AACAAAAGGT GTTTAAGGAT GAACTAAATA TTTTAATTGA	900
GCATTTATAT GGATAATCAT ACATTTATGTA AGCCCATATG TATTTACATC CAGAGTCATA	960
ATATTTAAA TAAACAATCA TGCAGAAACT TTTTTAGGG GTATACTATT GTTTAATAT	1020
55 CGTTGCCAAT TTGCTGACT TAAAATATGT GACATTTAA AATCAGGATT TTCCATATTN	1080
G	1081

60

(2) INFORMATION FOR SEQ ID NO: 22:

5	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 807 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:	
	GAATTCCGCA CGAGCTCCCTT CAGAAATGTC TTGGCTATTC TTGCTCTTTC CTCTTCTCTG	60
15	TAAATTTCAAG CATAAACTTA RTTTCCATAA TATATGACTG GAAATTTAC AGAAGAGTTA	120
	ATGTGTCTAA CTAGCAAACA CGAAGAAAAG CTCAGTGTAA GCAGTTAACT GAGGGAAATGC	180
	AAATCAAGAC CACAAGGAGA TAACAATTTC ACCCTATTGA CAAAAGTCA GAAGTCTAAT	240
20	AATACTAAGT GTTGGAGAGG ATATGCCCA GTATGATCTT ATCCACTGTT GGTGGAGATA	300
	TCAATTAGTA CAAACACTTT GAAAATAAG ARGGAATTCT ATAATATCTA ACATTTGCAT	360
25	ATATCCATTT ATCTCTCTAG ATCTAGATCT TAGCCCTCTC CACCCGTGAC TGTGTTCTTG	420
	GAAGGGGATC ATGAATGGTT TCCTTGCATT CTGCCTCTG ATTTGGTCA GCCAATGAGA	480
	GACCATGGCA AGACATTTGT GAGAAGGGTA GAGAGTCAGG TCAAGGTTCT TAGTGAGATC	540
30	AACTCTTCT CTGCCAGTTT GTTAACTGAA TTCTACTGAA AGCTAGAGCT CTGTTGAGTA	600
	ATCTTTAAA GCTGCAGCTA CCCTTTTGAG ATTAAGTAAT AGCTCCCTGT TTGTCCTTG	660
35	TTAGGGCTAG GGATGTTAA GGATCCTTGC CCTTGCTAGT CCTAGCATGT TTTGGTGTCC	720
	CATAATAGTT CTTTTTTAA ACTTCCTCA ATTAACACAAT TTGATCTTGT TCCTACCACT	780
	ACCCATTGCTG GTACAAACCTT AAACTGG	807
40		

(2) INFORMATION FOR SEQ ID NO: 23:

45	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 632 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:	
	GAATTCCGCA CGAGCTCTAAC ACCATAAAAGA AATAACAGCT GCATTCAAGA CCAGGATATG	60
55	TAAAATAATT TGTGTTAGTTT CAGCCACTTT TTAAAGTCAA TTTTACACC TGAAAGAAAG	120
	GCAATCCTGA CTCCATTGTT CTTTCGCCAA TAAGGAGATC GGGAAATTACA ATAATAAATA	180
	GAAGAAAGAA TGTTGCTTTT CCTCACTGTA ATTAATTAA TGGCTCTTGC GAAGATGAAT	240
60		

	TTTTGTGGTG ATTAAAATAG TCCCTTGCAC ATATTAGGTA CTCAGTAAGC ATTTGTGAAA	300
	TAGGGACTTT CTAGCCTTTA TTTGTGTTTA AGGAATCAGG GAATAAGTC AAAATGCCT	360
5	TTCAAGAAAT TTTTGAACT CTCTTCAC TAAGAAACTG TAAAGTCTTA TAAAAGAGAC	420
	ATTATTTATT TCTCTCAAGT ATTGCTTGCG AGGTGAATIG AAGGTTTTT TTTTATCAAC	480
10	AGTTGTTTTA TAAGATCGTT TGAGGACTAA AAGGGCTGAT TGTAATCACC TGAAACATGT	540
	TACCCAGCAA GACATTCTC ACCAGGTTGA AGTAAAAAAA ARAAATGAAG TGAGAATATC	600
	AAGCTTATGC AAGTTTGAAA TTNCAAACAA GA	632

15

(2) INFORMATION FOR SEQ ID NO: 24:

20	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1358 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:
----	-------------------------------------------

	GGCACGAGGA TAAATTGCAA GTATTAATCG GTCCCAACTT TAATATGGGA TAAAATAAC	60
30	AGTCAGTATG TGACCTCCTA AACAAATCCCT CTACTGAGCT GTGGAGGGGA GAAGGGAGGT	120
	CCTGGGGCCA GGACAGACAG GGCTATTTTC AGTAGTACAA CTTATATGCT ACTCTAAGAA	180
35	AACTCCAGAA AATGCRATTC TCTTCATACG AAGTCTTARA TACCCCTCATK ATTTRGATAA	240
	ATACATTTTC ARRCTAATA TGGAGACAGA AAGCTGCCTA GATTTATACC CACAAGTATT	300
	ATAAATTITAG AGAGTCTGAC CAGCCTCAAT TATTTCTCTT CGAAGTGGGA GAGAGAAATC	360
40	AAAAGTCAGA AATGGTGGRT AATCTCCAAG TCATATCCAT TTGGSTTTGR TCTACTACTT	420
	GTTTTTATGC TTTGTTTTGG RGRCAAGGRT GCCTGATGTT AAGGGRATTT CMTACMTGAA	480
45	ATAATGTGAC CAGACTGCCA TCTAGTCAAA AACCTATAAA ATGTTATTTA CTTTAATTCT	540
	GGGCTTAATTC AACAGAAGTY YYSGATAAAA RCTCTCCAAA CAATAATTAT GARCCTTAGT	600
	TTTTGTGTTT GTTTGGATA CAAAACAAAA CAGCTCTGTA GTGTTCTGT GAGGTTTATA	660
50	AATAGATTTT TTAACTACT TAATTTTCYG GTTTCYGCCY CTGKGTTTCY TGTACCTATA	720
	GAGGTAGCTC TTTTCAGTTA AGTAGAGAAA AGCTCTTCCC CTGGGTTGAA AATAATGCAG	780
55	TCCCGAGAGG CTACTTAACt CTACCTTTCT GGAGGTCATG GTAGCAATTG GAGATCTCCC	840
	AGGCATTCTA AGGGGAGCTA CTAAGAGCC CCAGATACTC AATTTACAC TAGAAATTG	900
	CTTCATCTAC TCTCTGTCT CAT CTGGGGAGRA AAGTATTATA ACTGACATTC AGTATGCACA	960
60	CAATAAGTGC ATAATAAAGA CCTATTGAGG GGATCCAAGG GAGTAAAATG GGTTTGCCCA	1020

	TAGGACTCCA TCAGGGTCCA CCAACACAGA CTTACAGCAA AAATTGGAAG GCTCTTTCT	1080
5	GCTGGATTCT GGGAACTCTGT GTTCTCTAGT GTGCCAGGGA GAGTTGGAAT CAAACACGT	1140
	AATATAATGT TTCTATTCAAG AGCCCCATTT TTTTGCCAAA TAAAGTAGCA CTGTCAAATA	1200
	ATAAAATCTTG TATTCACTTG GCCATGTATG TTATTATTG GATCTCTAAA ATATGCTTCA	1260
10	AATAATGCAC TGAAATAAGT GAGGTGATGA ATTTTGAAAT AATAACAGTT TATGATGGT	1320
	AGCTCCAAA TTTTTAAAAA AAAAAAAA AAACCTCGA	1358

15

(2) INFORMATION FOR SEQ ID NO: 25:

	(i) SEQUENCE CHARACTERISTICS:
20	(A) LENGTH: 1376 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

	CCCACCTTTA GCGAGCCAAC GAGAGAACAC CGCCTGCAGC TAGAACAGCC TGGTCAGGAG	60
30	CGTAACGGAG TGGTGGCCA ACCGTGAGAGG AAACCCGTGC GGGGCTGGC TTTCCTGTCC	120
	CCAAGCCGTT CTAGACGCGG GAAAAATGCT TTCTGAAAGC AGCTCCCTTT TGAAGGGTGT	180
	GATGCTTGGA ACCATTCTGT GTGCTTGTAT CACTATGCTA GGACACATTA GGATTGGTCA	240
35	TGGAAATAGA ATGCACCACC ATGAGCATCA TCACCTACAA CCTCCTAAC AAGAAGATAT	300
	CTTGAAAATT TCAGAGGATG AGCCCATGGA GCTCAGTAAG AGCTTTCGAG TATACTGTAT	360
40	TATCCTTGTAA AAACCCAAAG ATGTGAGTCT TTGGGCTGCA GTAAAGGAGA CTTGGACCAA	420
	ACACTGTGAC AAAGCAGAGT TCTTCAGTTC TGAAAATGTT AAAGTGTGTTG AGTCAATTAA	480
	TATGGACACA AATGACATGT GTTAAATGAT GAGAAAAGCT TACAAATACT CCTTTGAWAA	540
45	GTATAGAGAC CAATACAACG GGTCTTCCT TGACGCCCT ACTACGTTG CTATCATTGA	600
	AAACCTAAAG TATTTTTGT TAAAAAAGGA TCCATCACAG CCTTTCTATC TAGGCCACAC	660
50	TATAAAATCT GGAGACCTTG AATATGTGGG TATGGAAGGA GGAATTGTCT TAAGTGTAGA	720
	ATCAATGAAAG AAGCTTAACA GCCTCTCAA TATCCCAGAA AAGTGTCTTG AACAGGGAGG	780
	GATGATTTGG AAGATATCTG AAGATAAACAG GCTAGCAGTT TGCCTGAAAT ATGCTGGAGT	840
55	ATTTGCAGAA AATGCCAGAAG ATGCTGATGG AAAAGATGTA TTTAATACCA AATCTGTTGG	900
	GCTTCTATT AAAGAGGCAA TGACTTATCA CCCCAACCAG GTAGTAGAAG GCTGTTGTTTC	960
60	AGATATGGCT GTTACTTTTA ATGGACTGAC TCCAAATCAAG ATGCATGTGA TGATGTATGG	1020

5	GGTATACCGC CTPAGGGCAT TTGGCATAT TTTCAATGAT GCATTGGTT TCCTACCTCC	1080
	AAATGGTCT GACAATGACT GAGAAGTGGT AGAAAACCGT GAATATGATC TTTGTATAAGG	1140
10	ACGTGTGTTG TCATTATTTG TAGTAGTAAC TACATATCCA ATACAGCTGT ATGTTTCTTT	1200
	TTCTTTCTA ATTTGGTGGC ACTGGTATAA CCACACATTA AAGTCAGTAG TACATTTTA	1260
	AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	1320
	AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAA	1376
15	(2) INFORMATION FOR SEQ ID NO: 26:	
	(i) SEQUENCE CHARACTERISTICS:	
20	(A) LENGTH: 2923 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
25	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:	
	CTCCCTCCTCC GGGGCCCCCT CCTCCCCCTT TMACTGGTGC AGATGGCCAG CCTGCTATAC	60
	CACCAACCGCT TTCTGATACC ACCAAGCCCA AGTCCTCCTT GCCTGCCGTG AGCGATGCC	120
30	GTAGCGACCT GCTTTCAAGC ATCCGTCAAG GTPTTCAAGT GCGCAGGGTT GAKGAGCAGC	180
	GGGAACAAGA GAAGCGGGAT GTTGTGGCA ATGACGTGGC CACCATCTTG TCTCGTCGA	240
35	TTGCTGTTGA GTACAGTGAC TCAGAAGATG ACTCCTCTGA ATTTGATGAG GACGACTGGT	300
	CCGATTAACT CTTCTGCGCT GCTGCCACC TTCTTTCTT TTCCCTCCTA CCTGCCCTCT	360
	TTGATGCCAA CCCAACAGA CCCGTAGGG AGGAAAAGG AGGAAAAAAG TAATTTAAG	420
40	GGGCAAAGC TTCCCTGAA GCAACCAAAG ATATATCCAA GTGCTTCCTC CAAGTCACAA	480
	TGTATTTCTT CTCCCCATTTC TCAGGCCCTG TGGGCTCCT GAGGTCAGT AGCTGGGATG	540
45	TTCCCTCTTT CCTTCAGTG CCTGTTGCAT ATTGAAAGGA AGGAGAAATC CCAAAGCAGA	600
	TTCCTTGTAT CGGGTTCTTG TTGGAGATGG GCGTTCCCTT AGGAGCCATA TTCAACTACA	660
	GCCTTCTAAA ACCTGTGCCCC TCAGCCACTT CGAATGCCAG CCACCTCTG GTTCTAAAC	720
50	GGGGAGTGGT CTGAATGAAC ACAGCTGACC CCTTTCCCGC GCACTGAAAG GGCAGAGTAG	780
	GCGGAGGTC CAAGGCCAG ACTGCCACAC CCTCTGCCCT AATCAGCAGG GTGGGCCCTGC	840
55	CTTTGCTAA GCGATCTCTA TGCCCTGGAT GCGCTTTATT CCAGGAGGCA TCAAGCCTCT	900
	AAAGAATGTC TCACCTCCTC TGCCCAAAA TGATGCCCTT CTGTTAGCTG GTGTTGTTGC	960
	CTCCCTCCCA GGATCCCTT GGTGAGTATG GTGTTAGGA TGCACCAACCA CCACCTCTAG	1020
60	ATACCTTCAG GCAACACAGC CCAGTTTAA CCTCTAGTAT CCATGACCAA ACTATCCCTG	1080

	ACACATGAGG ACAGGGGCCT CTTCTGGCTG TCAGGAGCAA AGCCTGAAGA CTTGGAGCTG	1140
5	CAGGACTGGA AGAACAGTGG AGCCCCGIGG GTCTCACCCCT TTAAGGATGC TGAGGCCTAG	1200
	AGATGGGAAG TGACTTGCTC AAGGTACACAC AATTGGATAG TGACATAGCT AGACGCCAGA	1260
	GTTCTGATT CCAAGTCACC TGTGCTTCT GGGACCAAG AATGGGCACC TGCTGGAGTC	1320
10	CGGGCAGAGC TTTCTCAGTT GTATTGCTAC TCCAGACCTC ACCATAGGTT GGGGTCCCAG	1380
	TAGGAAGGCT CAGGGCTGTG GCCACCCCTG TCGGTGCTGC TCAGACCTTC ATAGCCCTCTC	1440
15	TTGTCATTCT TTGTTGCCCTT TTTTCTGTCA CCAGCCAACC ACATAGCCCTT GGGACCAGCC	1500
	TCTCTGGGGG ACCAGAAGTA GTGAGAGAAG GAAGGGGATA GGCAGCTTGTG ACAGGTGCTG	1560
	CTTCAATTTC CTCTGCAACT CCTCCCCCTT TTATTTCCCC AATTTAAACA AAGATTCTGC	1620
20	CAACTGTGGA AACTTCAGTC CCTCAGGCTG GCAGCCATGC CAGTACCTGC CTGGGGGTGG	1680
	GGGGTGCCTG GCAGCCATGA AGCAGGCTGA AAGGCAGAGG GGCTCCAGGT CCTGTTCCA	1740
25	GCTCCCCCTCA CTGGCACATGG TGAAGCTCGC TCCCTCCCTC CCTCCCCCTCC CGCTTTCCCC	1800
	AGAGCTAATA CACAGGTGCT ATTATTCAGA AAAAAACTGG TCAGCTCTAG CCAACAGTGA	1860
	AGGTTTCTTT TCTTCTGCCCT TNAACTATTG TGTAGCCTCT TATGCTGAAA TCGGCTCTG	1920
30	CTGGCTCTC CGGCTTTCAG AGCCCTGAAA CAAAGAGAAA CAGGATCTGT CCCTACCCAG	1980
	CACAGCAAAT GGTGTAGTA ATTGCCAAAG CCCTCATAAA GCCTCCGGC TTGAGGAGAG	2040
35	AGTGTATAGT CATGGGTTCT GCCTCTGTGC CCTTGCTGGC CGCTTCTCTT CTGGCTCTT	2100
	TCCTTGGAACT CACGGTGTGG GGACTGAGCC TGTAGGGGAC AGCATGCCGT CTTGCTGTGG	2160
	CCACTCCCAA GTGTGCCCTC TTCCCTCTTT ACACATCAGG TGTCTCTGCC ACAGGACTTG	2220
40	GCACTAAGCT CCATGCTGAG ACACCAGGCT ATGTGGGCC CCACCTTGT TCCCAGGCTG	2280
	CACCTTAAAGA GCGGAAGTGC TTTCATCAGA ACCCTAAAAT GGTGTTGAA GGCCCTGGG	2340
45	CCGCAGCCAG CAGTAGTTGG AGAGGCAGGC AGAGGGCAGT GGTCTCCCA AATAGGAGAC	2400
	CTGGGGCTG CCCAGGCAGG GTTTGGGCCT AATGGCTTGT ACTAAATTAC CCCCATCCTC	2460
	CTTGCCCCGA AAAGGGAGAG CTAGAGCCAC TCACTGTCA TCTGCTCTGA CCTTGAGGG	2520
50	GGCGGTGTTG GCCTGGCTTC TGGATGGAC TGAGTCCATC GTGGAAAGGG CTGGGGCAG	2580
	GAGGAGGTGG GGAGGGGCAC TGCCTGCGGA AGGTAGGATT AGATCATTAG CTCAGTGACC	2640
55	TCCTAGGGTT TCGATGTGCT ATGTTCTCAT CCTACAGTTG GTTTGGTAAT GATCTGCAAG	2700
	TCCCCGGAGAG CAACAGCACA GCTCTGCCCTG ACGCTCTCAT TAAAATCTAT GCAGCCAAGC	2760
	TGGCACTTT GTAGCAGCCG GCCTTGCGAA GCCTCCTCAG CTCGGGGGGC CGGGGACCCA	2820
60	GTGAGCCGNA GAKCSTCTGG GCTCCACTTA TGCAATGCA CCAAAAAAAA AAAAAAAA	2880

AAAAGGGGG CGCGCTCTANA AGGATTCCTC NAAGGGGCC AAG

2923

5

(2) INFORMATION FOR SEQ ID NO: 27:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 775 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GAACTAGTGN ATCCCCGGG CTGCAGGAAT TCGGCACGAG CCCRACCCSC	ACCACCA	60
GAATGCAGTT CCAGCTTAGG AAGCCACAAA CAAGCCACCC AGGAGGAACA	AAACACCGCC	120
20 ACCGTGGATT TTCCCAAAT TCCCTGGAAA GTAAGTCTCG CTCTTGCCAA	AGAAAAGTCT	180
GGCTTGGAGA GTCTCTGGAG CCCAGGATGC CAGCATGTGC CAATGACTGT	CACCTTCATC	240
25 TCTTCAAAAG AAAAGCCATA GCGGAGGACT GTCCCGCGAC CCCCGTGGAC	TGCGTCTAGG	300
TCATGTGATT CTGTTTCAT TTCTCATCCC ATCCAATTG TCCCTTTCTC	CTGTCATTTT	360
30 CTTCCCTCTGT GGTCCCTTCA AAGTTGTAT AATTTGTACT GAACTTCAA	ATGTGTCCCG	420
TTCTCCCCAG ACCACTCTAG CCACAGTATA TTGCAATAAA ATTACTTCCT	ATATTTGCAG	480
AAATTCCTTTT GGTGTAATTT TATTTTTTCC TCTCAATATA TATAATTGGA	CAAACGCTGG	540
35 CAAAAAGAAA AAAATGGTAA GCAAAAAACC CAAGATAAG TTTCGAGGAC	ATCAGGCCCTT	600
TTGAAATACA ATGTCAAATG ACACATTGTA CGKTTTCAA	AAATCCGCTA GACATGTCT	660
40 AAGTTTTAAC TGTAAATGCC AGGAAAGGAT ATCTTAAAT ATTCTAAACT	TGTGTAACAA	720
AGGAATAATT AACTGTAATA GTTTTCAAT AAATCGAGTT GGGTGTTC	ACCGT	775

45

(2) INFORMATION FOR SEQ ID NO: 28:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 534 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GAATTGGCA CGAGCAAGGG TGGAACCTGA GTCTGCTTGT CTGTTTGCCC	CATGACAGCC	60
CAGGGGTGGT GGSCTCACCC CACCTCCAGG CAMCCACAAG AATATAAAAT	CTTGTACAAR	120
60 GATGTCGATA TTACTATTGS CATTCCAAG TGCACCTGCA CCTGTAGTAT	CAGGTGGTTT	180

	GCAGCCTTGG CTGCATAGCT GCATATGAGA ATCACCTGGG AAGCTTTAA AAATCCCAGT	240
5	ATCCCCACCT CTTCCCCAGT TACAGTGGAG TCTTGCGGGT GGTGGGGAC ATCAATTATT	300
	TTTGAAGCT CCMAAGTAAT TCTGGTGTGC AGTGGGGTGA CCAGCTGTCC CAGGGAMTC	360
	CTTTAAAAAA TAATATCCCG GGCACATGAC AGGCCAATTG CCCTAATGCA ACCAAGGTTA	420
10	AGAACTACTG GTTTAATGGG AAAATATTTC TTTCCTGTC TTGAATAATA CTGGTTTAT	480
	TAAACTCCNG AATCCCATT CTTTCCCTGCA CAAATTTTT AAAGGCNAAA AAAA	534

15

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

20	(A) LENGTH: 1827 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

	NNNCNGCACGA GCNCGGTCCT GTCCCGTCAG CGTCCCCCA GCCAGCTCCT TGCACCCCTC	60
30	GGGGCCGAGG CCCTCCCTGG TGGTCCCCGC GCAGCCATGG CTCAGCACCTT CTCCCTGGCC	120
	GCCTGCGACG TGGTCGGATT CGACCTGGAC CACACTCTGT GTCGCTACAA CCTGCCGAG	180
	AGCGCCCCGC TCATTTATAA TAGCTTGTGC CAGTTCTAG TTAAGGAGAA AGGGTACGAT	240
35	AAGGAATTGC TCAATGTGAC CCCAGAGGAT TGGGATTCT GTTGCAAAGG TTTGGCATTG	300
	GATCTAGAAG ATGGGAACCTT CCTTAAACCTT GCAAATAATG GCACTGTTCT CAGGGCAAGC	360
40	CATGGCACCA AGATGATGAC TCCAGAGGTG CTGGCAGAGG CATATGGCA GAAAGAGTGG	420
	AAGCACTTCT TGTCGGACAC TGGAATGGCT TGCGCTCAG GAAAGTATTA CTTTACGAC	480
	AACTACTTTG ACCTGCCAGG AGCTCTTCTG TGTGCCAGGG TGGTGGACTA TTTAACAAAA	540
45	CTGAACAATG GTCAAAAAAC ATTTGATTTC TGGAAGGATA TAGTTGCTGC TATACAACAC	600
	AATTATAAAA TGTCAAGCTTT TAAGGAAAAC TGTGGAATAT ATTTTCCAGA AATAAAAAGA	660
50	GATCCAGGCA GATATTTACA TAGTTGTCTT GAATCTGTGA AAAATGGCT TCGACAGCTA	720
	AAGAATGCTG GGAAAATTCT TCTGTTAATT ACCAGTTCTC ACAGTGATTA CTGTAGACTT	780
	CTCTGCGAAT ATATTCTTGG GAATGATTTC ACAGACCTTT TTGACATTGT GATTACAAT	840
55	GCATTTGAAGC CTGGTTCTT CTCCCCACTTA CCAAGTCAGA GACCTTTCCG GACACTCGAG	900
	AATGATGAGG AGCAGGAGGC ACTGCCATCT CTGGATAAAC CTGGCTGGTA CTCCCCAAGGG	960
60	AACGCTGTCC ACCTCTATGA ACTTCTGAAG AAAATGACTG GCAAAACCTGA ACCCAAGGTT	1020

	GTTTATTTTG GTGACAGCAT GCATTCAGAT ATTTTCCCAG CTCGTCACTA TAGTAATTGG	1080
	GAGACAGTCC TCATCCTGGA AGAACTCAGA GGGGATGAAG GCACGAGGAG TCAGAGGCCT	1140
5	GAGGAGTCAG AGCCTCTAGA GAAGAAAGGA AAATATGAGG GACCAAAAGC AAAACCTTTA	1200
	AATACTTCAT CTAAAAAAATG GGGCTCTTTT TTTATTGATT CAGTTTTGGG ACTGGAAAAT	1260
10	ACAGAAGACT CCTTGGTTTA TACATGGCT TGTAAAGAGAA TCAGTACTTA CAGCACTATT	1320
	GCAATTCCAA GTATTGAAGC AATCGCAGAA TTACCTCTGG ACTACAAATT TACAAGATT	1380
	TCTMCAAGCA ATTCAAAAAC ACCTGGCTAC TATCCAAATC CTCCACTGGT CITATCAAGT	1440
15	GATGAGACAC TGATATCCAA ATAAGTTGTC TTTACTGAAA AATGAAGTGA AGACCCATAT	1500
	ATGCAGTTAA AAAAAAGTTA ATTTTCAAAA AATACTGTAA AAGACTTTAA GGAACAAGTT	1560
20	TTATTGACCA ATAAGTTGAT ATTTGTCCAT AGGTCTCCCT TCTATAAATC ATCTTGATGT	1620
	TTAACAACTC TTATTATATT AAAATCTCAG TATCCTAAAA CTTAGGAACC TTATTGGATA	1680
	TTTTCTATTA CAGTAGTTT GTGGTTGGGA TTCACCCGGG GGGGCCACAC ACTCACACGG	1740
25	CACAGTCAC TCTTACACA TATGGCCNCG GTCCCCGTGGG GTTCTCNAAG GTGTGGTCC	1800
	CTTGGGGCCT NTGGGCTTG GCCCTTT	1827

30

(2) INFORMATION FOR SEQ ID NO: 30:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1479 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

	GGCAGGAGGG CGGGTGGCAT CAGCAGAGGG GCACCAAGCA AAGGGTGTGG CTACCTCACT	60
	GCTGGTCCCC AGGCCCCGGGA GGTGGGGAGC ACACACAGTG CCTTGGGTAC CCAGNTGGGT	120
45	GTTCTCCCGC TGCAAGAGGAG ACRCGAGCCT GGGTCTGTC CTTCACCTCT GGCGGCCTTC	180
	TCTACATCGC CTGGTGAAC GTGCTCCCTG ACCTCTTGGAA AGAAGAGGAC CGGTGGCGCT	240
50	CCCTGCAGCA GCTGCTCTG CTCTGTGCGG GCATCGTGGT AATGGTGTGTC TTCTCCCTCT	300
	TCGTGGATTA ACTTTCCCTG ATGCGACGC CCCTGCCCTC TGCAAGCAATA AGATGCTCGG	360
	ATTCACTCTG TGACCGATA TGTGAGAGGC AGAGAGGGCG AGTGGCTGCG AGAGAGAATG	420
55	AGCCTCCCGC CAGACAGGAG GGAGGTGGT GTGGATGTAT GTGGTGTGCA CATGTGGCCA	480
	GAGGTGTGTC CGGGAGACCG ACACGTGAT CCCTGTGCTG GGTCGGGGC CCAGTGTAGC	540
60	GCCTGTCCCC AGCCATGCTG TGGTTACCTC TCCCTGCCGC CCTGTCACCT TCACCTCCTG	600

	GAGTAAGCAG CGAGGAAGAG CACCACTGGT CCCAAGCAGA GGCCTTGCCC TGCTGGGACC	660
5	CGGGAGTGA GAGCAGCCCA AGGATCCCAG GGTCCAGGGA ACTCCAGAGC TGCCCACCTC	720
	CCACTGCCCC CTCAGCACAC ACACAGTCCC CAGGGGGCCT AGGGGCCAAG GCTGGGGCGG	780
	CTTTGGTCCC TTTCTCTGGC CCTTCTCTCC CCACITCTAA GCCAAAGAAA GGAGAGGCAG	840
10	GTGCTCTGT ACCCCAGCCC CACTCAGCAC TGACAGTCCC CAGCTCTAG TAGTGAGCTG	900
	GGAGGCCTT CCTAAGACCC TTTCTCTAGG GCTGCCCTGG GAGCTCATTC CTGGCCAACA	960
15	CGCCCTGGCA GCACCAAGCAG CTCTTGCCAC CTCCAGCTGC CAAACAGCAG CCTGCCGGGC	1020
	AGGGAGCAGC CCCAGGCCAG AGAGGGCTCC CGGTCCAGCT CAGGGATGCT CCTGCCAGCA	1080
	CAGGGGCCAG GGACTCCTGG ACCAGGCACA TAGTGAGCCC GGGCAGCCCT GCCCAGCTCA	1140
20	GGCCCTTTTC CTTCCTCATT GAGGTTGGGG TAGGTGGGGG CGGTGAGGGC TCCACGTTGT	1200
	CAGGGCTCAG GAATGTGCTC CGGCAGAGTG CTGAAGCCAT AATCCCAAC CATTTCCTT	1260
25	GGCTGACGCC CAGGTACTCA GCTGGCCAC TCCACAGCCA GGCCTGCCCT GCCCTTCACC	1320
	GTGGATGTTT TCAAGAAGTGG CCATCGAGAG GTCTGGATGG TTTTATAGCA ACTTTGCTGT	1380
	GATTCCGTTT GTATCTGTAA ATATTGTTC TATAGATAAG ATACAAATAA ATATTATCCA	1440
30	CATAAAAAAA AAAAAAAA AACTTGGGGG GGGGNCCCG	1479

35 (2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 987 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

45	GGCACGAGCG CAATCGCGTT TCCGGAGAGA CCTGGCTGCT GTGTCCCGCG GCTTGCGCTC	60
	CGTAGTGGAC TCCGGGGGCC TTGGCAGAT GCAGGCCTGG GGTAGTCTCC TTTCTGGACT	120
50	GAGAAGAGAA GAATGGAGAA GCCCCTCTTC CCATTAAGTGC CTTTGCATTG GTTTGGCTTT	180
	GGCTACACAG CACTGGTTGT TTCTGGTGGG ATCGTTGGCT ATGTAAAAAC AGGCAGCGTG	240
	CGTCCCTGG CTGCAGGGCT GCTCTTCGGC AGTCTAGCCG GCCTGGGTGC TTACCAAGCTG	300
55	TATCAGGATC CAAGGAACGT TTGGGTTTC CTAGCCGCTA CATCTGTAC TTTTGTGTG	360
	GTTATGGGAA TGAGATCTA CTACTATGGA AAATTCATGC CTGTAGGTTT AATTGCAAGT	420
60	GGCAGTTTGC TGATGGCCGC CAAAGTTGGA GTTCGTATGT TGATGACATC TGATTAACAG	480

	AAGTCATGTT CCAGCTTGGGA CTCATGAAGG ATTAAAAATC TGCATCTTCC ACTATTTTCA	540
	ATGTATTAAG AGAAAATAAGT GCAGCATTTT TGCATCTGAC ATTTTACCTA AAAAAAAA	600
5	GACACCAAAAT TTGGCGGAGG GGTGGAAAAT CAGTTGTTAC CATTATAACC CTACAGAGGT	660
	GGTGACCATG TAACATGAGC TTATTGAGAC CATCATAGAG ATCGATTCTT GTATATTGAT	720
10	TTTATCTCTT TCTGTATCTA TAGGTAATC TCAAGGGTAA AATGTTAGGT GTGACATIG	780
	AGAACCCCTGA AACCCCCATTC CCTGCTCAGA GGAACAGTGT GAAAAAAAAT CTCTTGAGAG	840
	ATTTAGAATA TCTTTCTTT TGCTCATCTT AGACCACAGA CTGACTTTGA AATTATGITA	900
15	AGTGAATAT CAATGAAAAT AAAGTTACT ATAAATAAWA AAAAAAAA AAAAAAAA	960
	AAAAAAA AAAAAAAA ANANAAA	987

20

(2) INFORMATION FOR SEQ ID NO: 32:

	(i) SEQUENCE CHARACTERISTICS:
25	(A) LENGTH: 2933 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

	TCTACCTCCG AGTAGTATTA GACTGTAAAC ACAGTAATAT AGNCGCCATC ATTCTGTGAAG	60
	GGGTTTCTTT TGGGGGACAG AGGATCAGAT GTTGAGAGTT TGGACAAACT CATGAAAACC	120
35	AAAAATATAC CTGAAGCTCA CCAAGATGCA TTTAAAATC GTTTGCGGGA AGGTTTCTG	180
	AAAGCTCAAG CACTCACACA AAAAACCAAT GATTCCTAA GGCGAACCCG TCTGATTCTC	240
40	TTCTTCTGC TGCTATTCCG CATTATGGA CTTCTAAAA ACCCATTCTT ATCTGTCCGC	300
	TTCCGGACAA CAACAGGGCT TGATTCTGCA GTAGATCTG TCCAGATGAA AAATGTCAACC	360
	TTTGAACATG TAAAGGGGT GGAGGAACCT AAACAAGAAT TACAGGAAGT TGTGAATT	420
45	TGAAAAATC CACAAAATT TACTATCTT GGAGGTAAAC TTCCAAAAGG AATTCTTTTA	480
	GTGGGACCCC CAGGGACTGG AAAGACACTT CTTGCCGAG CTGTGGCGGG AGAAGCTGAT	540
50	GTTCCTTTT ATTATGCTTC TGGATCCGAA TTTGATGAGA TGTGTTGTGG TGTTGGAGCC	600
	AGCCGTATCA GAAATCTTT TAGGAAAGCA AAGGCGAATG CTCCCTGTGT TATATTTATT	660
	GATGAATTAG ATTCTGTGG TGGGAAGAGA ATTGAATCTC CAATGCATCC ATATTCAAGG	720
55	CAGACCATAA ATCAACTTCT TGCTGAAATG GATGGTTTA AACCCAATGA AGGAGTTATC	780
	ATAATAGGAG CCACAAACTT CCCAGAGGCA TTAGATAATG CCTTAATACG TCCTGGTGT	840
60	TTTGACATGC AAGTTACAGT TCCAAGGCCA GATGTAAAG GTCGAACAGA AATTTGAAA	900

	TGGTATCTCA	ATAAAATAAA	GTTTGATCAW	TCCGTTGATC	CAGAAATTAT	AGCTCGAGGT	960
5	ACTGTTGGCT	TTTCCGGAGC	AGAGTTGGAG	AATCTTGTGA	ACCAGGCTGC	ATTAAAAGCA	1020
	GCTGTTGATG	GAAAAGAAAT	GGTTACCATG	AAGGAGCTGG	GAGTTTTCCA	AAGACAAAAT	1080
	TCTAATGGGG	CCTGAAAGAA	GAAGTGTGGA	AATTGATAAC	AAAAACAAAAA	CCATCACAGC	1140
10	ATATCATGAA	TCTGGTCATG	CCATTATTGC	ATATTACACA	AAAGATGCAA	TGCCTATCAA	1200
	CAAAGCTACA	ATCATGCCAC	GGGGGCAAC	ACTTGGNACA	TGIGTCCCTG	TTACCTGAGA	1260
15	ATGACAGATG	GAATGAAACT	AGAGCCCAGC	TGCTTGCACA	AATGGATGTT	AGTATGGGAG	1320
	GAAGAGTGGC	AGAGGAGCTT	ATATTTGGAA	CCGACCTATAT	TACAAACAGGT	GCTTCCAGTG	1380
	ATTTTGATAA	TGCCACTAAA	ATAGCAAAGS	GGATGGTTAC	CAAATTTGGA	ATGAGTGAAA	1440
20	AGCTTGGAGT	TATGACCTAC	AGTGTACAG	GGAAACTAAG	TCCAGAAACC	CAATCTGCCA	1500
	TCGAACAAGA	AATAAGAATC	CTTCTAAGGG	ACTCATATGA	ACGAGCAAA	CATATCTTGA	1560
25	AAACTCATGC	AAAGGAGCAT	AAGAATCTCG	CAGAAGCTTT	ATTGACCTAT	GAGACTTTGG	1620
	ATGCCAAAGA	GATTCAAATT	TTTCTTGAGG	GGAAAAAGTT	GGAAGTGAGA	TGATAACTCT	1680
	CTTGATATGG	ATGCTTGCTG	TTTTTATTGC	AAGAATAYAA	GTAGCATTGC	AGTACTCTAC	1740
30	TTTTACAACG	CTTTCCCCCTC	ATTCTTGATG	TGGTGTAAATT	GAAGGGTGTG	AAATGCTTGT	1800
	TCAATCATTT	GTACACATTAA	TCCAGTTGG	GTAACTCTCA	TTATGACACC	TATTGCAAAT	1860
35	TAGCATCCC	TGCAAATAT	ATTTGAAAAA	AATAAAGAAC	TATCAGGATT	GAACACAGCT	1920
	CTTTGAGGA	ATGCAATTAA	GTATTAAAGT	TGAAAGTAAT	TAATGATTTT	ATGTTGGTT	1980
	ACTCTACTAG	ATTTGATAAA	ATTTGTGCCT	TTAGCCTCT	ATATACATCA	GTGGAAACTT	2040
40	AAGATGCAGT	AATTATGTTTC	CAGATTGACC	ATGAATAAAA	TATTTTTAA	TCTAAATGTA	2100
	GAGAAGTTGG	GATTAAGAAC	AGTCTCGGAA	ACACAGGCC	AGGGAAATATA	GCCTTTGGC	2160
45	ATGGTGCCAT	GGCTCACATC	TGTAATCCC	GCACCTTTGG	AGGCTGAGGC	GGGTGGATTG	2220
	CTTGAGGCCA	GGAGTTCGAG	ACCAGCCTGG	CCAACGTGGT	GAAACGCTGT	YTCTACTAAA	2280
	ATACAAAAAA	ATAGGGCTGG	GGCGGGTTGC	TCACGCCCTGT	AATCCCAGCA	CTTTTCAGAG	2340
50	GCCAAGGCGG	GCAAATCACC	TGAGGTCAAG	AGTTTGAGAC	CAGCCTGCC	AACATGGTGA	2400
	AACCCCATCT	CTACTAAACA	TGAAAAAATT	ACCTGGCAT	GGTGGCAGGT	GCTTATAATC	2460
55	CCAGCTACTC	TGGGGGCCAA	GGCAGGAGAA	TTGCTTGAGC	CTGGGAGATG	GAGGTTGCAG	2520
	TGAGCTGAGA	TCATGCCACT	GCACCTCCAGC	CTGGGCAACA	GAGCAAGACT	CTGCCTCAA	2580
	AAAAAATTAA	AATAAATTAA	AATACAAAAA	AAAATAGCCA	GGTGTGGGGT	GCATGCCTGG	2640
60	AATCCCAGCT	ACTTGAGAGG	CTGAGGCACG	AGAATTGCTT	GAACCCAGGA	GGTGGAGGTT	2700

	GCAGTGAGCC AAGATCACAG GAGCCACTGC ACTCCAGCCT GGGTGACAGA GTGAGACTCT	2760
5	GTCTAAAM AAAATTAAAT AAATTATTAT AACCTTTCAG AAATGCTGIG TGCATTTC	2820
	TGTTCTTTT TTAGCATTA CTGTCACCTCT CCCTAATGAA ATGTACTTC GAGAAGCACT	2880
	ATTTTGTTAA ATAAATACAT AACCTCAAAA AAAAAAAA AAAAAAAACT CGA	2933

10

(2) INFORMATION FOR SEQ ID NO: 33:

15	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1366 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:
----	-------------------------------------------

	GGGAATACCT ATTCTCCCTT ACCGTGTGTC TTTTCCCCCT GGAATTGAGC CAGCAAGTTC	60
25	TGGCATGGC AGGTGTTCT GAAATATCAG TGTGTTTTT TTTCTTCTT TTGTTTCCT	120
	TGTTTTCGTC TTCTATTTT CCTAAGCAGG CAACTCCAAA AAGAGATTG TTTGTGCAGG	180
30	AGTCAGGAAA AGGGAAGAGG AATACTGAAA GCTGGGAGTA GGGCAGGACA GAAGACGGGG	240
	AGGAGTCTAT TTTCATTGTC TAAGTGTGTA ACTTCCACCA ATGCCAAAGT CACGGACATG	300
	TGTGCAGTTG GATGTCGAG TTAGAGCAGC CCCAAGGGCC TGTAAACCTGA ATAGCAGGCA	360
35	CTCACCCAGC TGATAACTCA AGTTCCAAAT GGACCACAGC TGAGTTGTAG GGGATGTGTC	420
	TGTGTGTGTA CGCGTGCCTT TGAGATTCCCT GGAACAGATT TCCTCTGAGA TCTCAACAGG	480
40	CTTTTCATT ATCATTGGGG AGCTATGGTT TCTCTTATTTC CACAAGGCC ATTTCTTCCT	540
	TTTGAGATGT GCAAGGAGAT GACTCCATCC ATGACTTGGC TTTACACTCT CCCTCCCTGG	600
	CTTTTATCA TCAGTGCAGR AGARATTCTT GCTCGTTCTT CAAACAATCT CATTGAGCT	660
45	TTATAAAGAT TATTGGARTT TAAATAATAT TCATATCTAT GGCCTAGAAC AATGTTCCTC	720
	AAGTATGCGT CAGAAATCATG AGTGGTAGAG GGAGGATTAT AATGTAGTTT CCTACATTTC	780
50	TACCTCCAC CACCTGGAG TCTGCATTTC AACGTACTTC TGTYTGAGGA TCAGAYTTG	840
	GGAAAGCGTG GGCTTGAGAT GTTTCTKGA CATTGATTTA TGTTGAGACC AGACCAAGAA	900
	GCAGATGGAT GGACATGATC AGTTCACTAA CATGTTCTT TCTTAGGGTC AAATTGGAGG	960
55	AGGCTCTAGA GAAGCACTGT CCAATAGAAA TATAATGCCA ACAATATATG TWAATTTAAG	1020
	TCTTCTATTG GTGCATTAA AAAGTAAAAG AAGGCTGAGT GGCTGGCAT GGCTCCTCGT	1080
60	GCCTGTAAATC CCAGCACTTT GGGAGGCCGG GGTGGGCAGA TCACCTGAGG TCAGGAGTTC	1140

	GAGACCAGCC TGCCCAACAT GGTGAAACCC CATAATNTACT AAAAATACAA AAAATTAACC	1200
	GGCATAGTG GCAGGTGCCT GTAAATCCCAG CTACTCGGGA GGCTGAGGCA GGAGAATCGC	1260
5	TTGAAACCTGG GAGGCAGAGA CTGCAGTGAG CTGAGATCGT GCCACTACAC TCCAGCCTGG	1320
	GTGATGAGCG AAACCTCGTC TCAAAAAAAA AAAAAAAA ACTCGA	1366

10

(2) INFORMATION FOR SEQ ID NO: 34:

	(i) SEQUENCE CHARACTERISTICS:
15	(A) LENGTH: 667 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

	ATTTTCGGCA CAGGCCGGAA GCTACCTATC TGGTAGGGAG CTCCCCCAGC ACCGAAGACT	60
	GGGATGACTT CTGCRCTGAC CCAGGGCTG GAGCGAATCC CAGACCAGCT CGGCTACCTG	120
25	GTACTGAGTG AAGGTGCACT GCTGGCGTCA TCTGGGGACC TGGAGAAATGA TGAGCAGGCA	180
	GCCAGTGCCA TCTCTGAGCT GGTCAAGCACA GCCTGGGGTT TCCGGCTGCA CGCGGGCATG	240
30	AATGTGCCCT TCAAGCGCCT GTCTGTGGTC TTTGGAGAAC ACACACTGCT GGTGACGGTG	300
	TCAGGACAGA GGGTGTGGTG GGTGAAGAGG CAGAACCGAG GTGGGGAGCC CATTGATGTC	360
	TGAGGCTGCC GGAGGGCGAG GGTGGAGAA GCGGATTGGG TCTGGGCCT CTGTGATGAG	420
35	GCAGGCACAN CTGTCGGTCT TGGCTTGCTG CTAGAACTAG GGCCTTCTGC TCGCCCACCT	480
	CCACACCTA CCTGGACGGG CCCAGGCTTG GGGACTCTGA CCTGTGTTAA GGAGAACAAAG	540
40	GGCAAGGAGA CCTCCCTTGT TGCTCCCTCA CTCCCTAATA AACATGAGTC TGATGTTCTC	600
	CARMMAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA	660
45	AAAAANN	667

50 (2) INFORMATION FOR SEQ ID NO: 35:

	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1710 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
55	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

60	GGCACCGAGCC AGAGCCAGGCT GCTAGGCCTG GGGGCCACAC TGCCCCCTGGG TGCTACACCC	60
----	----------------------------------------------------------------------	----

	AGTGTGCTGG GTCACTGGGA ACTTCCTGAA GTGGTGTAC CTGAACGTGGG CCCCCAAGGA	120
	TGGGGTGGGG GCAGTACCGC AGGAAGAGGA GCAGCCCTG TGAAGATGAA GAGCTGCCAG	180
5	AGGCTCTGTG ATTGGCTGCG GCACGATGAC CGCGCAGGG ATTGGCTGCT TCGGGCGGG	240
	GGCCCGGGCC CGGGGGACAG AATCCGCCCC CGAACCTTCA AAGAGGGTAC CCCCCCGCAG	300
10	GAGNTGGCAG ACCTTAGGAG GTGCGACAGA CCGCGGGGC AAACGGACTG GGGCCAAGAG	360
	CCGGGAGCGC GGGGCCAAAG GCACCAGGGC CGCCCGAGGG CGCGCGCGAG CACGGCTTG	420
	GGGGTTCTGC GGGCCTTCGG GTGCGCGTCT CGCCTCTAGC CATGGGTCC GCAGCGTTGG	480
15	AGATCCTGGG CCTGGTGCTG TGCCTGGTGG GCTGGGGGG TCTGATCCTG GCGTGGGGC	540
	TGCCCATGTG GCAGGTGACC GCCTTCCTGG ACCACAACAT CGTGACGGCG CAGACCACCT	600
20	GGAAGGGCT GTGGATGTG TGCGTGGTGC AGAGCACNNG GCACATGCAG TGCAAAGTGT	660
	ACGACTCGGT GCTGGCTCTG AGCACCGAGG TCCAGGCGGC GCGGGCGCTC ACCGTGAGCG	720
	CCGTGCTGCT GGGGTTGTTGTTGCG TGACCCCTGGC GGGCGCGAG TGCACCACCT	780
25	GCGTGGCCCC CGGCGCGGCC AAGGCGCGTG TGGCCCTCAC GGGAGGGTG CTCTACCTGT	840
	TTTGGGGCT GCTGGCGCTC GTGCCACTCT GCTGGTTCCG CAACATTGTC GTCCGGAGT	900
30	TTTACGGACCC GTCTGTGCCCG GTGTGCCAGA AGTACGAGCT GGGCGCANGC TGTACATCGG	960
	CTGGCGGGCC ACCCGCGCTGC TCATGGTAGG CGGCTGCCTC TTGTGCTGCG GCGCTGGGT	1020
	CTGCACCGGC CGTCCCGACC TCAGCTTCCC CGTGAAGTAC TCAGCGCGC GGCGGCCAC	1080
35	GGCCACCGGC GACTACGACA AGAAGAACTA CGTCTGAGGG CGCTGGGCAC GGCGGGGCC	1140
	CTCCCTGOCAG CCACGGCTGC GAGGGTTGG ATAAGCCTGG GGAKCCCCGC ATGGACCGCG	1200
40	GCTTCGGCGG GGTACCGCGG CGCGCAGGCT CCTCGGAACG TCCGGCTCTG CGCCCCGACG	1260
	CGGCTCCCTGG ATCCGCTCTT GCCTGCGGCC GCAGCTGACC TTCTCCTGCC ACTAGCCCG	1320
	CCCTGCCCTT AACAGACGGA ATGAAGTTTC CTTTTCTGTG CGCGCGCGCTG TTTCCATAGG	1380
45	CAGAGCGGGT GTCAAGACTGA GGATTTGCT TCCCCCTCAA GACGCTGGGG GTCTGGCTG	1440
	CTGCCCTTACT TCCCGAGGGC TCCGTGTGAC TTCCGGAGGG CGGATGCAGA GCCCAGGGCC	1500
50	CCACCCGGAA GATGTGTACA GCTGGCTTTT ACTCCCATCGG CAGGCCCCAG CCCAGGGACC	1560
	AGTGACTTGG CCTGGACCTC CGGGCTCCTAC TCCAGCATCT CCCAGGGCAA GGCTTGTGGG	1620
	CACCGGAGCT TGAGAGAGGG CGGGAGTGGG AAGGCTAAGA ATCTGCTTAG TAAATGGTTT	1680
55	GAACCTCTCAA AAAAAAAAAA AAAAAAAAAA	1710

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 1096 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

10	GGCCAGTGGG CAGGGTCACA GGGCAAGGTC CGCGGGGCGG CTGGGTGCGG CGACTTCCGT	60
	GCTCCCGGGG AGCGGGCGGA GAGCGGGGGC CGCACTGGGG AGTGTGGGCT GGGCCCGAGA	120
15	TGTCAATGTTG CCTGTCTTCTT GGACCGTGGT TCGTACCTAT GCTCCCTATG TCACATTCCTC	180
	TGTTGCCCTTC GTGGTCCGGG CTGTGGGTTA CCACCTGGAA TGGTTCATCA GGGAAAGGA	240
	CCCCCAGCCC GTGGAGGAGG AAAAGAGCAT CTCAGAGGCC CGGGAGGATC GCAAGCTGGA	300
20	TGAGCTCTA GGCAGGACC ACACCGCAGGT GGTGAGCCTT AAGGACAAGC TAGAATTTC	360
	CCCGAAAGCT GTGCTGAACA GAAACCGCCC AGAGAAGAAAT TAATGGAGGA CACAGGGCCC	420
25	TATGGTCCTA CTGTGGGTGG TGACTTGTCC TGCTACCATG TTGACAGAGC CCCAGAACCC	480
	ACATCTAATT GGCTTTGTTG CTTATTCTGG CCCTTCCCAC ACCACACAGC CACACAAATA	540
	CTGGCTGCTC CTTGATGGCC AGGCAGACCC AGCAGCAGCC GAGGGGCCAG TGAAGAGGAA	600
30	GGCCGCATCT GTTGTGTGGT GGCCACAAGC ACTCAGGCAT CTGAGTTAC TGGTGCCTG	660
	CTGGGAGGAG AGTTATGAGA TGAACATTGG CTGTCAATCT CTGTGGCAG GCGGTTGGC	720
35	CTCTAGTGGG AATGGCTGGG ATTTGGCGT TGCCCTTCTGG AGGGATACCT GCAATGTCTAG	780
	TTCCAGTCTG CACTGGAAAG AATTCAAATA TGCACCTGGC TCCCTTCAC TTTTGGCCCT	840
	ATCCCTTGTG CTCATTCTTA CTGAAATCTG TCTTGTCAAG TCAGGAATGG GATTCCTTCA	900
40	GGAAGGAAAG CACTTTCTG TTCTGGGAAG CCCAGACTGT TCACCTTGGG GCAGGGACGA	960
	ACATGTCCCT CGTGAATTTC CTTGAAAACA GTCAACCATCT TCTACCCCCCA TCACGTATA	1020
45	GTGAAAAACC TGATTAAGT GGTATCTGAG AACCAAAAAA AAAAAAAA AAAAAAAA	1080
	AAAAANGGGG GGNCCC	1096

50

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 2279 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

60

	GGTGGCCAAG GGGCTCAGCT CGCAGCGCAT GCCCCGCGCAC AGGTTCTGTC TGGCCGTGGG	60
	CAGCGCCGTC TTTAATGCCA TGTTCACACGG GGGMATTGCCA ACAACATCCA CGGAGATTGA	120
5	GCTGCCGAC GTRGAACCCG CGCCCTTCCT CGCACTGCTC AAGTTTCCTCT ACTCGGACGA	180
	GGTGAGATT GGCCCGGAGA CGGTGATGAC CACGSTATAC ACCGCCAAGA AGTACCGGGT	240
10	GCCAGCGCTC GAGGCCATT GCGTGGAGTT CCTGAAGAAG AACCTGCGAG CCGACAACGC	300
	CTTCATGCTG CTCACGCAGG CGCGACTCTT CGATGAACCG CAGCTGGCCA GCCTGTGCCT	360
	GGAGAACATC GACAAAAACA CTGCAGACGC CATCACCGCG GAGGGCTTCA CCGACATTGA	420
15	CCTGGACACG CTGGTGGCTG TCCCTGGAGCG CGACACACTG GGCATCCGTG AGGTGGGGCT	480
	GTTCAATGCC GTTGTCCGCT GGTCCGAGGC CGAGTGTCAAG CGGCAGCAGC TGCAGGTGAC	540
20	GCCAGAGAAC AGGGGGAAGG TTCTGGCAA GGCCCTGGGC CTCAATTGCT TCCCGCTCAT	600
	GACCATCGAG GAGTTGCTG CAGGTCCCCC ACAGTCGGGC ATCCCTGGTGG ACCGGAGGTT	660
	GGTCAGCCTC TTCTGCACCT CACCGTCAAC CCCAAGGCCAC GAGTGGAGTT CATTGACCGG	720
25	CCCCGCTGCT GCCTGCGTGG GAAGGAGTGC AGCATCAACC GCTTCCAGCA GTGGAGAGT	780
	CGCTGGGCT ACAGSGGGAC CAGTGACCGC ATCAGGTTCT CAGTCAACAA GGCATCTTC	840
30	GTGGTGGGAT TTGGGCTGTA TGGATCCCTC CACGGGCCA CCGACTACCA AGTGAACATC	900
	CAGATTATTTC ACACCGATAG CAACACCGTC TTGGGCCAGA ACGACACGGG CTTCAGCTGC	960
	GACGGCTCAG CCAGCACCTT CGCGTCTATG TTCAAGGAGC CGGTGGAGGT GCTGCCAAC	1020
35	GTCAACTACA CGGCCCTGTC CACGCTCAAG GGCCCAGACT CCCACTACGG CACCAAAGGC	1080
	CTGCCAAGG TGACACACGA GTGCCCAACC ACGGGCGCCA AGACCTGCCT CACCTTTGCT	1140
40	TACGGGGCGG GGAACAACAA TGGCACATCC GTGGAGGAGC GCCAGATCCC CGAGGTCACTC	1200
	TTCTACACCT AGGCTGCCCG ACACCGACAC CGCCCTCCCT CCGTGGGGAT AGCCGCAGCC	1260
	CCAGGCCATC ATCTGCTGCT GGGYCCCCC CACCAAGCGG TGCCAGGCC AGTGTCCCCC	1320
45	AGGCCGTCTG TCCACTCCAT GCCACCTTTC TCAGCATCAG GACGGGGTGTG CCCTGTGTTTC	1380
	ACCACGAGTK TGGCTGCTGG ATCAGGGCAG CGGGGGAGGT GGCCAGGCCA GTGGCCAGGC	1440
	CCTGTGGAGA CAATCCCTCA GGACTAGGGA CAGGGCTGTC CCGGCCCTGGG CCAGGGGCCA	1500
50	CGGACCCGCA GCTCAGGGCG CCTGCCCAAGC TCGTCTGCCG GCGGTGCGCC GCGGGCGTCC	1560
	CTCGCGTCTC TTCACTGCAC ATTGCAATGC ATTTCGGAATT CCCATTTCCTC TGCTAGGAGC	1620
55	CAGCTGGGT GGCGCTGCTC CCAGAGCCGT GGGTCCCAGA CCTTGCGTTC CTTTTGTTC	1680
	TGTCCGTTA TCAGGACACG GGCCCCACCT GTCACTGTC CGAGGCCACC CAAGCCCAGC	1740
60	CTGGGGGGCG TTCCCCACTGC CTGGATGCCG CCTTGAGTTTC TGCCGACGCCA GGATTCACTG	1800

	TGGGGACGGC CCCTGCCGGA TAGGCCTAGC CCTGGCCCAG GTGGTGAGCG GTTGCAGTG	1860
	TCCGTTCTCA TCCACCTGAT GGGCCAGAT AAAGGCCCCC GCTGTCCAGC CTCCCTGGAC	1920
5	GGCCCTCGCG GTCCCTGCAG CCCAAGATGG GACTCAGACC CTGTGCCCCA GAGCTCCCC	1980
	GGCGAGAAT GGGGCCAG CGGGCCCGA CGGGGTCCAG GAGCACTGCT CGCCTGTACA	2040
10	TACTGTTGCC CTAGCCCACC TGGTGCCGTG GGAGCCACCC CCAGGTGCTG GGGCACAGCC	2100
	CCTCCCCACT CGGCCACGC CCCCACCCAC CCCCGTGTGT TCTGCCCTGT GACTCCTGGA	2160
	ACCTGCGTCC TCCCCAAAGC CATGGGAGGG GTGTCCTCCT CAGACCATGC CCCCAGATGA	2220
15	TTTTTTTAAA TAAAGAAACA AATGCACCTG CAAAACAAAA AAAAAAAA AAAACTCGA	2279

20 (2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 745 base pairs
- (B) TYPE: nucleic acid
- 25 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

30	GTACAGGACT GAGAAGCAGA TAACAAGAGT GACGCTCACA GGGCTGGGCT GACGCTAAC	60
	GGAGGCAGTG TGTGGCTCGA AGATTCTTGA ACCCACAGCA GCAGCTGGG CCACCCATC	120
35	CTGCCACAG CTCCAGCCCT GAGACGACGA GGAGGAGAGT CGACTTTGCC TCTTGCCCAA	180
	GGGACCATGC CCAGGTGCCG GTGGCTCTCC CTGATCCTCC TCACCATTCC CCTGGCCCTG	240
	GTGGCCAGGA AAGACCCAAA AAAGAAATGAG ACGGGGGTGC TGAGGAAATT AAAACCCGTC	300
40	AATGCCCTCA ATGCCAACG TGGAAGCAGT GTYYGTGGTT TTGCCATGCA AGAATACAAC	360
	AAAGAGAGCG AGGACAGTA TGTCTTCTTG GTGGTCAAGA CACTGCAAGC CCAGCTTCAG	420
45	GTCACAAATC TTCTGGAATA CCTTATTGAT GTAGAAATTG CCCCCAGCGA TTGCAGAAAG	480
	CCTTTAAGCA CTAATGAAAT CGCGCCATTIC AAGARAACTC CAAGCTGAAA AGGAAATTAA	540
	GCTGCAGCTT TTGGTAGGA GCACCTCCCT GGAATGGTGA ATTCACTGTG ATGGAGAAAA	600
50	AGTGTGAAGA TGCTTAATGG TGTGTTGAGG CATCCCTCCA ACCTCTGTGA CTACTTTATC	660
	CATGAAAATG AAGCAATGGT CAGGTGGGAG GCTCTTCCCA ATGTGCTTTC TTCAAAAAAA	720
55	AAAAAAA AAAAAAAA CTCGA	745

60 (2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1718 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 5 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

10	CCCCATAGGC AGGAGGGCCC CGGGCAGCAC ATCCTGTCIG CTTGTTGCTG CTGCAGAGTT	60
	CTGTCCTTGC ATTTGGTGCGC CTCAGGCCAG GCTGCACTGC TGGGACCTGG GCCATGTC	120
	CCCACCCAC CGCCCTCCTG GGCCTAGTGC TCTGCCTGGC CCAGACCATC CACACCCAGG	180
15	AGGAAGATCT GCCCAGACCC TCCATCTCGG CTGAGCCAGG CACCGTGATC CCCCTGGGA	240
	GCCATGIGAC TTTCGIGIGC CGGGGCCCCG TTGGGGTTCA AACATTCCGC CTGGAGAGGG	300
20	AGAGTAGATC CACATACAAT GATACTGAAG ATGIGTCTCA AGCTAGTCCA TCTGAGTCAG	360
	AGGCCAGATT CCGCATTGAC TCAGTAAGTG AAGGAAATGC CGGGCCTTAT CGCTGCATCT	420
	ATTATAAGCC CCCTAAATGG TCTGAGCAGA GTGACTACTG GAGCTGCTGG TGAAAGAAC	480
25	CTCTGGAGGC CSGGACTCCC CGGACACAGA GCGGGCTCC TCAGCTGGAC CCACGGAGAG	540
	GCCGTCGGAC AACAGTCACA ATGAGCATGC ACCTGCTTCC CAAGGCCTGA AAGCTGAGCA	600
30	TCTGTATATT CTCATGGGG TCTCAGTGGT CTTCTCTTC TGTCTCCCTC TCCCTGGCT	660
	CTTCGCTTC CATGCCAGA ATCAGATAAA GCAGGGCCC CCCAGAACCA AGGACGGAGGA	720
	GCAGAAGCCA CAGCAGAGGC CTGACCTGGC TGTGATGTT CTAGAGAGGA CAGCAGACAA	780
35	GGCCACAGTC AATGGACTTC CTGAGAAGGA CAGAGAGACG GACACCTGG CCCTGGCTGC	840
	AGGGAGITCC CAGGAGGTGA CGTATGCTCA GCTGGACCAC TGGGCCCTCA CACAGAGGAC	900
40	AGCCCGGGCT GTGTCCCCAC AGTCCACAAA GCCCCATGGCC GAGTCCATCA CGTATGCAGC	960
	CGTTGCCAGA CACTGACCCC ATACCCACCT GGCCCTCTGCA CCTGAGGGTA GAAAGTCACT	1020
	CTAGGAAAAG CCTGAAGCAG CCATTTGGAA GGCTTCCCTGT TGGATTCTC TTCATCTAGA	1080
45	AAGCCAGCCA CCCAGCTGTC CTGGAGACAA GAGCTGGAGA CTGGAGGTTT CTAACCAGCA	1140
	TCCAGAAGGT TCGTTAGCCA GGTGGTCCCT TCTACAATCG AGCAGCTCT TGGACAGACT	1200
50	GTTTCTCAGT TATTTCCAGA GACCCAGCTA CAGTTCCCTG CCTGTTCTA GAGACCCAGC	1260
	TTTATTCACC TGACTGTTTC CAGAGACCCA GCTAAAGTCA CCTGCCCTGT CTAAAGGCC	1320
	AGCTACAGCC AATCAGCCGA TTTCCTGAGC AGTGTGATGCCA CCTCCAAGCT TGTCTTAGGT	1380
55	GTCTGCTGTG AACCTCCAGT GACCCAGAG ACTTTGCTGT ATTAACTGCTC CCTGCTGACC	1440
	CTAAAGACCT TCCTAGAAGT CAAGAGCTAG CCTTGAGACT GTGCTATACA CACACAGCTG	1500
60	AGAGCCAAGC CCAGTTCTCT GGGTTGTGCT TTACTCCACG CATCAATAAA TAATTTGAA	1560

	GGCCTCACAT CTGGCAGCCC CAGGCCTGGT CCTGGGTGCA TAGGTCTCTC GGACCCACTC	1620
	TCTGCTTCA CAGTTGTTCA AAGCTGAGTG AGGGAAACAG GACCTACGAA AAAAAAAA	1680
5	AAAAAAATCG AGGGGGGCC CGTACCCAAT CGCCTGTA	1718

10 (2) INFORMATION FOR SEQ ID NO: 40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1966 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

20	GTCGGCCTG CAGGTGACCA CTAGTGGATC CAAAGAATTG GGCACGGAGCT GGGGAGCGGG	60
	ACTSGAGAAT ACTGCCCAGT TACTCTAGCG CGCCAGGCCG AACCGCAGCT TCTTGGCTTA	120
25	GGTACTTCTA CTCACAGCGG CCGATTCCGA GGCAACTCC AGCAATGGCT TTTGCAAATC	180
	TGCGGAAAGT GCTCATCAGT GACAGCCTGG ACCCTTGCTG CCGGAAGATC TTGCAAGATG	240
	GAGGGCTGCA GGTGGTGGAA AAGCAGAACCC TTAGCAAAGA GGAGCTGATA CGGGACTGCA	300
30	GGACTGTGAA GGCTTATTG TTCGCTCTGC CACCAAGGTG ACCGCTGATG TCATCAACGC	360
	AGCTGAGAAA CTCCAGGTGG TGGCAGGGC TGGCACAGGT GTGGACAATG TGGATCTGGA	420
35	GGCCGCAACA AGGAAGGGCA TCTTGGTTAT GAACACCCCC AATGGGAACA GCCTCAGTGC	480
	CGCAGAACTC ACTTGTGGAA TGATCATGTG CCTGGCCAGG CAGATTCCCC AGGCAGCGGC	540
	TTCGATGAAG GACGGAAAT GGGAGCGGAA GAACCTCATG GGAACAGAGC TGAATGGAAA	600
40	GACCTGGGA ATTCTTGGCC TGGCAGGAT TGGGAGAGAG GTAGCTACCC GGATGCAGTC	660
	CTTGGGATG AAGACTATAG GGTATGACCC CATCATTTCC CCAGAGGTCT CGGCCTCCCT	720
45	TGGTGTTCAG CAGCTGCCCT TGGAGGAGAT CTGGCTCTC TGTTGATTCA TCACTGTGCA	780
	CACTCCTCTC CTGCCCCCTCA CGACAGGCTT GCTGAATGAC AACACCTTTG CCCAGTGCAA	840
	GAAGGGGGTG CGTGTGGTGA ACTGTGCCCG TGGAGGGATC GTGGACGAAG GCGCCCTGCT	900
50	CCGGGCCCCG CAGCTCTGGCC AGTGTGCCCG GGCTGCACTG GACGTGTTA CGGAAGAGCC	960
	GCCACGGGAC CGGGCCTTGG TGGACCATGA GAATGTCACTC AGCTGTCCCC ACCTGGGTGC	1020
55	CAGCACCAAG GAGGCTCAGA GCGCTGTGG GGAGGAAATT GCTGTTCAGT TCGTGGACAT	1080
	GGTGAAGGGG AAATCTCTCA CGGGGGTTGT GAATGCCCCAG GCCCTTACCA GTGCCTCTC	1140
	TCCACACACC AAGCTTGGA TTGGCTCTGC AGAAGCTCTG GGGACACTGA TGCGAGCCTG	1200
60	GGCTGGGTCC CCCAAAGGGA CCATCCAGGT GATAACACAG GGAACATCCC TGAAGAATGC	1260

	TGGGAACCTGC CTAAGCCCCG CAGTCATTGT CGGCCTCCCTG AAAGAGGCCTT CCAAGCAGGC	1320
5	GGATGTGAAC TTGGTGAACG CTAAGCTGCT GGTGAAAGAG GCTGGCCTCA ATGTCACCAC	1380
	CTCCCCACAGC CCTGCTGCAC CAGGGGAGCA AGGCTTCGGG GAATGCCTCC TGGCCGTGGC	1440
	CCTGGCAGGC GCCCCTTACCC AGGCTGTGGG CTTGGTCAA GGCACATACRC CTGTACTGCA	1500
10	GGGGCTCAAT GGAGCTGTCT TCAGGCCAGA AGTGCCTCTC CGCAGGGACC TGCCCCGTCT	1560
	CCTATTCGGG ACTCAGACCT CTGACCCCTGC AATGCTGCCT ACCATGATTC GCCTCCGTGGC	1620
15	AGAGGCAGGC GTGCGGGCTGC TGTCCTACCA GACTTCACTG GTGTCAGATG GGGAGACCTG	1680
	GCACGTCAATG GGCAATCTCTC CCTTGCTGCC CAGCCTGGAA GCGTGGAAAGC AGCATGTGAC	1740
	TGAAGCCCTTC CAGTTCCACT TCTAACCTTG GAGCTCACTG GTCCCTGCCCT CTGGGGCTTT	1800
20	TCTGAAGAAA CCCACCCACT GTGATCAATA GGGAGAGAAA ATCCACATTC TTGGGCTGAA	1860
	CGCGGGCCCTC TGACACTGCT TACACTGCAC TCTGACCCCTG TAGTACAGCA ATAACCGTCT	1920
25	AATAAAGAGC CTACCCCCAA AAAAAAAA AAAAAAAA ACTCGA	1966

30	(2) INFORMATION FOR SEQ ID NO: 41:	
	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 972 base pairs	
	(B) TYPE: nucleic acid	
35	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:	
40	GGCACGAGCC AAGTGGTCCC CCAGACAAGG CTCAGGATGT CCACATCCAC TGCATCCTGG	60
	ACCTGTGCA GGTGAAGATG TCCCGACCCA CGCATACTCC TCTTTCGCCCT GCCACCATT	120
	CTCCAACCAT CACAGTAGCA GTCTTCTTCG CTGTGTTCTG CGCCGCCGCCGCCGCCACCG	180
45	CCGTGTGCGC CGTGGCTGCT GCAACCACCA GCAGCGGSCG CAGAACTASA GACAAATCCC	240
	CCATAGCCAC TCACTCTTCC GTAACCCACA TCGCAGCCAA AAGATGTCAC AACTACACCG	300
	AGTGCCTTTC TTGATCAGG ARGACCCGGA TTCTTACCTG GARGARGARG ACAACCTGCC	360
50	CTTCCCGTAT CCCAAGTACC CACGTGGGG CTGGGGGGGG TTTTATCAGA GAGCGGGCCT	420
	GCCTCCAATG TGCGCTGCTG GGGCCACCAAG GGTGTATCCT GGCCAGTCTG CCACCAACCT	480
55	CTCTCTACCT GTCACCTGAG CTGGCTGCCA TGCCCCAAGG TGTAGAGGCC AGGTCTGAGC	540
	TGAGGCTCTG CCCCGCTGGC GTCTCTGAC TACCTCTGCC TCCCTCACGG TGTTGGACGA	600
60	GGCCCTCCAT CAAAGGACCC CAGCTCCAAG CTCAGTGTG TGCCCCCAATT CCTCCCCAGCC	660

	CTGGCCCCAA	GTCCAGGCTG	CGGACCCCTGC	CCCTCCCCCG	ACCATGTGTTG	TCCCACTCAG	720
	CGGAAATCCA	GGGGCCAATG	CCAACATACCA	GGTGTACGAC	AGCCTGGAGC	TGAAGGGCA	780
5	GOTGCAGAAG	AGCAGAGCCA	GGTCAGCTC	ACTGCCACCG	GCTTCCACCT	CCACCTTGAG	840
	GCCCTTYCTG	CACAGGAGCC	AGACCGAGAA	ACTCAACTGA	CCACCGAGGCG	GATGTGGGGT	900
10	GTGGGGCAGG	GCATGGAGGG	AGAGGAATAA	AGAGAAACAG	AGTCCAGGAA	AAAAAAAAAA	960
	AAAAAAA	ACTC	GA				972

15

(2) INFORMATION FOR SEQ ID NO: 42:

	(i) SEQUENCE CHARACTERISTICS:						
	(A) LENGTH: 1536 base pairs						
20	(B) TYPE: nucleic acid						
	(C) STRANDEDNESS: double						
	(D) TOPOLOGY: linear						
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:						
25	GGCACAGGCC	AACITAGTTT	GAGTTCTTCT	TCTGGACTCT	GTATGTCCTT	GTGTGTACCC	60
	TATGCCGTT	ACAGTCGTA	CTCTCTCTGT	GARATTGGCT	GTCTAATCCA	GGTGGATCAG	120
30	GAGGTGCTTT	GTGGTTTTTT	TGCAAAGAAA	TGAAGTCTGG	CAAGCAAACA	ATGATTAAC	180
	ATGTTTGCAT	TCCGTACTTG	TCTTTGGCG	AAATGCAAAG	GTGGGTGTGC	ATTCTTGAAT	240
35	TCAAAGAAA	TCTCTTCAA	ATCCCTCAT	CCCTTGTGTC	TCTTCTAAAT	ACTCTCTTTC	300
	TAGATATCTT	GCACCCCCAA	AACTCCCTCA	GCCCCCATGG	CAGCTTTCT	CTCTCCCTC	360
	TCTCTTCCC	GCCTCTCCCT	GTCTCTCAC	TTCAGCCTT	CCTCTTCTTT	AGATCTTTAT	420
40	TATGTAGATA	AAAACCCCTC	CAACCTCCCT	AGCCTTCTCT	CCATTGCACTC	CCCTACCCGA	480
	ATTATCCCTCA	AGAAAGAGCC	CAAGGTCCGA	CACAGGGATC	AGAAATCCCTC	CTCCCTTASA	540
45	AGCSCAGGGG	TGAGGGAGTT	CAGGAATAATT	CATAACTGG	TAATCCTTGT	CCCTGTACAA	600
	GTCACCTCT	TGTATCAGGA	CCCTGTAC	TATTTACAGA	CTATTTCCA	TCTCTCCTAA	660
	TGCAATTGCT	CAAAGGGCAC	TTTAAGNATA	ATCATTATCC	ATTGATGTTT	TTTGGAGGCT	720
50	TTTATTCCCT	CCAATAAGTT	CTGCCGAATA	CTGGCCGCTG	GCTCTATTG	TTAAACAAATG	780
	GAGGGCTTTG	TTCCGCTTTT	TTTTTTTTTT	TTWTTCTAA	CCTGAGCTTT	CTGCCACCC	840
55	TTAGTATGGG	GCCAAAGGG	AGATTTTAT	GCCACCCCTT	TTGGTGAGAA	GAGTCACCTC	900
	CTGATTAGTG	TTTGGGCTGA	AAATGGGTCC	CCCTTTGGGA	AGAAACATGG	GTGCAGTGTA	960
	CTTCTGTGT	CACAGGATTA	ACAGCTCTG	CCCCACTCCC	AAGGAGGCCAG	CTCYTCGGGG	1020
60	CAGTTCYTCT	TTGAGAATTT	CATGGTCATT	AAGAAGCAGG	YTCCCAGGGA	CCCCAGAGTG	1080

	GGAAACCTTTG ACTGAAGTCA CCACAGTGGG TGTAAGATAA ACATAAGAGA CTTTTCTCAG	1140
5	GGAAAGATTTG GAAAGAAGAA AAAGAGTAAA AAGTTCACAT GGACCATGGA GTGTTNTGGA	1200
	AAAGGGCCA GAAAGGGAAAG CTGTGGCTAA GAAGATAAAC TGCCTGATTC CAGAGACCCA	1260
	GGAGAGGGG TGAAATCTCT TTGTCGGTC ACATTTCTCW WTAATGATKY TCCACATGTA	1320
10	CAAAGCTAGC CAGTTTACCA AGTGCTTCCA CACACATTGC TTCATTCTGT GTCTCTTAAG	1380
	CAGATTGACT CCTTGGAAAA GCCTCACGTC TGGCATTCTG CACCTGCCA TCACCAGTT	1440
15	GGCCTTGGTC TGCTTGGCTG GTTGGGTCTC CCCATGGTGA GCTCCCATGG TATCTCTCT	1500
	TCACCTTTAT ATCACTCATT AGACACCGGT GACAAC	1536

20

(2) INFORMATION FOR SEQ ID NO: 43:

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 2541 base pairs	
25	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:	
30	AATTGGCAC GAGGTTCCCTG GCCAACCTGC TGCTGGAGGA GGATAACAAG TTTTGTGCAG	60
	ATTGCCAGTC TAAAGGGCGG CGATGGGCCT CTIGGAACAT TGGTGTGTTTC ATCTGCATT	120
35	GATGTGCTSG AATCCACAGG AATCTGGGG TGACATATTC CAGGGTAAAG TCAGTTAAC	180
	TCGACCAGTG GACTCAAGTA CAGATTCAAGT GCATGCAAGW GATGGGAAAT GGAAAGGCAA	240
40	ACCGACTTTA TGAAGCCTAT CTTCCGTGAGA CCTTTCCGGCG ACCTCAGATA GACCCAGCTG	300
	TTGAAGGATT TATTCGAGAC AAATATGAGA AGAAGAAATA CATGGACCGA AGTCTGGAC	360
	ATCAAATGCCT TTAGGAAAGA AAAAGATGAC AAGTGGAAAA GAGGGACCGA ACCAGTTCCA	420
45	GAAAAAAAT TGGAACCTGT TGTTTTTGAG AAGGTGAAAA TGCCACAGAA AAAAGAAGAC	480
	CCACAGCTAC CTCGGAAAAG CTCCCCGAAA TCCACAGCGC CTGTCATGGA TTTGTTGGC	540
50	CTTGATGCTC CTGTGGCCTG CTCCATTGCA AATAGTAAGA CCAGCAATAC CCTAGAGAAG	600
	GATTAGATC TGTTGGCTC TGTTCCATCC CCTTCTTCTT CGGGTTCCAG AAAGGTGTA	660
	GGTTCCATGC CAACTGCAGG GAGTGCAGGC TCTGTTCCCTG AAAATCTGAA CCTGTTCCG	720
55	GAGCCAGGG ACAAATCAGA AGAAATAGGC AAGAAACAGC TCTCTAAAGA CTCCATTCTT	780
	TCACTGTATG GATCCCAGAC GCYTCAAATG CCTACTCAAG CAATGTTCAT GGCTCCCGCT	840
60	CAGATGGCAT ATCCCCACAGC CTACCCCCAGC TTCCCCGGGG TTACACCTCC TAACAGCATA	900

	ATGGGGAGCA TGATGCCCTCC ACCAGTAGGC ATGGTTGCTC AGCCAGGAGC TTCTGGGATG	960
	GTTGCCCTTCA TGCCCATGCC TGCAGGCTAT ATGGGTGGCA TGCAGGCATC AATGATGGGT	1020
5	GTGCCGAATG GAATGATGAC CACCCAGCAG GCTGGCTACA TGGCAGGCAT GGCAGCTATG	1080
	CCCCAGACTG TGTATGGGT CCAGCCAGCT CAGCAGCTGC AATGGAACCT TACTCAGATG	1140
10	ACCCAGCAGA TGGCTGGGAT GAACTTCTAT GGAGCCAATG GCATGATGAA CTATGGACAG	1200
	TCAATGAGTG GCGGAAATGG ACAGGCAGCA AATCAGACTC TCAGTCCCTCA GATGTGGAAA	1260
	TAAAAAACAAA ACACCTGTAT GGCTGCCATT CTCTTCAGCC CTCGCTCTCC CCTTTCCACA	1320
15	GCCTCCACCC CTGACCCCCA TCCCTTTTC CTACCTCTCT GTTTGGTTA GAAATTGCTC	1380
	AATAAGTCAT TTGGGGTTTG GCATCCTGCC CAGCCACTTC CCAAACATGA AGACCTCTCT	1440
20	GTGCTTTAT GTTGTACATG CCCCATAGCC ATCCCAACGT CCTCCCCAGT CCTCTCCTGG	1500
	CACCAGCACC TTAGAAGTTG TTGGCAGAAG GCACTTAAAC TGTGGGAGAA GTGTCACAC	1560
	CTTTGAGTCC CTTCCCTCAA GGTTAAAGCT CCTGTCAGAC TCTCAGAAGG GTCTGTGGGT	1620
25	GTGTATATT AGCCAAACAG GGGAAAGCTT AGAGGTCTT CTATATGTGT TAATAAGCTG	1680
	TTTCTAAGTG TTTAAATTG AAAAGCATCA TGTCTCATG ATTTATGGGA ATGAAGCAAG	1740
30	TACTGAAATC AAATTAAATA CTCCTGGGT CCTGGGTCAAG TTGACCCCTA GCCCTGGGGT	1800
	GAGGCAAGCC CCCTCCTATG AGGATGAGCA AAAATACTAC TCTCTTCGCC CTGAGTTGCT	1860
	TTCCTGGATCT GGGGCTTCAG GACTTGCTGC TTCAGTCAGC CTTTATTAGC ACCAAAGACT	1920
35	TTATGAAGAT CCCACACACA GACACACATC CCTTCCCGCC TCCCCCTGC CTTCACTAGG	1980
	ATCTGGCTCC GTGGCTGGAG GACCAACCCC TATAGTGGGA ATGCAGAGCT TAACGTGTAC	2040
40	TGCTGTGTG TGTCGGTGAG TGTCGTGTG TGATGAGTG TGTCGTCAGC CTCACCCCT	2100
	CTCCCCATCT GCTCTGGTA TTTTTGTTTG TGTTTAGTT TAGGTTTACA ACAGAGAGGA	2160
	ATTAATTAT CAGCAGCCTA AAACGTGTGT GTTTTCTTA TGGTTAAAA AACGCCATGT	2220
45	CATTGATAAC TCCCTTTCTC CCTTCCCTTC TCCCGGTCTG CTGATCACTC TTTCATGCCT	2280
	GTGTATCCAG GGTGCTCTGT TTCCCCACCG TTCCCAGGTG TACGAGGCCAG AGGGCCGGGA	2340
50	CAGCTTCCCT CTCAGTCATT GTTCACCCCA CTTGAAAATT CAGACAAGAA AACTTTGCTT	2400
	AAAAGATTTC ATGTGTGGGA ACCACAGTTG CTGGCTGCCT TTCTCTGTG TATGTGTAAA	2460
	TTCCTTAATA AATATTGCAG GGAAGGACAA AAAAAAAA AAAAAAAA AAAAAAAA	2520
55	AAAAAAAAA AAAAAACTCG A	2541

60 (2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 2418 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

10	CCCAACGCGTC CGGCCAACGCG TCCGCCAACG CGTCCGCCA CGCGTCCGGG ACTCAGCGAA	60
	GGGTGGGCCG CGCCGAGGCC TCCGCCGCT GGCGGGTTTC CGCGGAGTC CGCCCGGCTC	120
15	CGCTCTGCCG CGGCCGCCGC TCATGGGCAG AGTCCGCCGG GCGGGCCGGC ATTAAACTGA	180
	AGAAAAGATG TCCCTGTACG ATGACCTAGG AGTGGAGACC AGTGAATCAA AAACAGAAAGG	240
	CTGGTCCAAA AACTTCAAAC TTCTGCAGTC TCAGCTTCAG GTGAAGAAGG CAGCTCTCAC	300
20	TCAGGCAAAG AGCCAAAGGA CGAAACAAAG TACAGTCCTC GCCCCAGTC TTGACCTGAA	360
	GCGAGGTGGC TCCTCAGATG ACCGGCAAAT TGTGGACACT CCACCGCATG TAGCAGCTGG	420
25	GCTGAAGGAT CCTGTTCCA GTGGGGTTTC TGCAAGGGAA GTTCTGATTC CCTTAGCTGA	480
	CGAATATGAC CCTATGTTTC CTAATGATTA TGAGAAAGTA GTGAAGCCCG CAAAGAGAGG	540
	AACGACAGAG ACAGCGGGAG TGGAAAAGAC AAAAGGAAAT AGAAGAAAGG GAAAAAAGGC	600
30	GTAAAGACAG ACATGAAGCA AGTGGGTTTG CAAGGAGACC AGATCCAGAT TCTGATGAAG	660
	ATGAAGATTA TGAGCGAGAG AGGAGGAAA GAAGTATGGG CGGACTGCCA TTGCCCCACC	720
35	CACTCTCTG GTAGAGAAAG ACAAAAGAGTT ACCCGAGAT TTTCCTTATG AAGAGGACTC	780
	AAGACCTCGA TCACAGTCTT CCAAAGCAGC CATTCCCTCC CGAGTGTACG AGGAACAAGA	840
	CAGACCGAGA TCTCCAACCG GACCTAGCA CTCCCTCTC GCTAACATGG GGGCACGGT	900
40	GCGCACAAG ATCATGCAGA AGTACGGCTT CCGGGAGGGC CAGGGCTGG GGAAGCATGA	960
	GCAGGGCTG ACCACTGCCT TGTCACTGGA GAAGACCAGC AAGCGTGGCG GCRAGATCAT	1020
45	CGTGGCGAC CCCACAGAGA AAGGTGTGTC CCCAGGGAAAG CGTGTGACTA GAGGGAAAGG	1080
	ACTGGCCCCA TCCATATCAG ACATGGCCAG TCTTGATCCT CATGTGTACG CAGGGGGACA	1140
	ATGAGGGCTG TGGCCAGAGG GAGAGGGCTG GCCCTGCCAT CACTAGAACAA CAGGCCGTCC	1200
50	TGTTCATATG ATGCACTGCC ACTTCGGTTT TGTGAAACCA GGAATCCTGA GGTCATCTT	1260
	TATTTTTTCA GAACAGACGT AGAGAGATGA AGGCTGTGG AGGAAAAGAT CGTGAGAGAC	1320
55	TTGGGCAGAA AATGAGTAGT CCTCAGGAAG AAATCTTGGT TATGTGTGTTA GAGCATGAAG	1380
	GACAGAGCCA TATAGTGTGG CAGTGAATAT ACCTGCTATC TCCATCTCAG AGGTGCTTC	1440
	TACTTTTCCC TTTTGCCCTT TCAGTATAGA TGTGATTCT GATTCTCTTA CAGATTGTTT	1500
60	GCTTGCAG ATCTGATGTT ATGTTGCAGT CTCTTGGTAA ATGATGCCAA GTTGGTGTGTTT	1560

	TATTTTCATT TAATTTTAC AGTCTGTTCT GTGTTGAGGG AATTCAAGGA AGAGACAAAC	1620
5	ATATGTTAGC ATTTTAATCA GGGAAATTAAG TTTGAGTCAG CCTAGCTGAA CTTCCCTTGCA	1680
	TAAAGAAAAGA AGAAAACCTTT TCTGGCAGCC CCGTTCATGC ACAGCTTAGG GATACATCAC	1740
	GACCTGACA GATGCATCCA AGAAGTCAGA TTCAAATCCG CTGACTGAAA TACTTAAGTG	1800
10	TCCTACTAAA GTGGTCTTAC TAAGGAACAT GGTTGGTGGG GGAGAGGTGG ATGAAGACTT	1860
	GGNAAGTTGA AACCAAGGAA GAATGTGAAA AATATGGCAA AGTTGGAAAA TGTGTGATAT	1920
15	TTGAAATTCC TGGTGCCCT GATGATGAAG CAGTACGGAT ATTTTGTGAA TTTGAGAGAG	1980
	TTGAATCAGC AATTAAAGCG GTTGTGACT TGAATGGAG GTATTTGGT GGACGGGTGG	2040
	TAAAAGCATG TTTCTACAAT TTGGACAAAT TCAGGGTCTT GGATTTGGCA GAACAAGTTT	2100
20	GATTTTAAGA ACTAGACAC GAGTCATCTC CGGTGATCCT TAAATGAACG GCAGGCTGAG	2160
	AAAAGAAGGA AAAAGGTACAC AGCCTCCATG GCTGTTGCAT ACCAAGACTC TTGGAAGGAC	2220
25	TTCTAAGATA TATGTTGATT GATCCCTTTT TTATTTGTG GTTTTTTAAT ATAGTATAAA	2280
	AATCCTTTTA AAAAAACAAC AATCTGTGIG CCTCTCTGGT TGTTTCTCTT TTTTATTATT	2340
	ACTCCTGAGT TGATGACATT TTTGTTAGA TTTCATGGTA ATTCTCAAGT GCTTCAATGA	2400
30	TGCAGCAATT CTTGCACT	2418

35 (2) INFORMATION FOR SEQ ID NO: 45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1337 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

45	TCGACCCACG CGTCCGGAGC GACCTCTCTG CTCCGCTCGT CTGTTGGTT CCGGAGGTGCG	60
	CTGGCGCGGT GGGAAATGCT GCGCGCGCGG GCGCGGGGCA CTGGGGCCCT TTTGCTGAGG	120
50	GGCTCTCTAC TGGCTCTCTGG CGCGCTCCG CGCGCGCGCT CCTCTGGATT GCCCCGAAAC	180
	ACCGTGGTAC TGTTCGTGCC GCAGCAGGAG GCCTGGGTGG TGGAGCCAAT GGGCCGATTC	240
	CACCGGATCC TGGAGCCTGG TTTGAACATC CTCATCCCTG TGTAGACCG GATCCGATAT	300
55	GTGCAGAGTC TCAAGGAAAT TGTCAATCAAC GTGCCTGAGC AGTCGGCTGT GACTCTCGAC	360
	AATGTAACTC TGCAAATCGA TGGAGTCCTT TACCTGGCAGCA TCATGGACCC TTACAAGGCA	420
60	ACCTACGGTG TGGAGGACCC TGAGTATGCC GTCACCCAGC TAGCTCAAAC AACCATGAGA	480

	TCAGAGCTCG GCAAACCTCTC TCTGGACAAA GTCCTCCGGG AACGGGAGTC CCTGAATGCC	540
	ACCATTTGTCG ATGCCATCAA CCAAGCTGCT GACTGCTGGG GTATCCGCTG CCTCCGTTAT	600
5	GAGATCAAGG ATATCCATGT GCCACCCCGG GTGAAAGAGT CTATGCAGAT GCAGGTGGAG	660
	GCAGAGCGGC GGAAACGGGC CACAGTTCTA GAGTCTGAGG GGACCCGAGA GTGCCCATC	720
10	AATGTGGCAG AAGGGAAGAA ACAGGCCAG ATCCCTGGCT CCGAAGCAGA AAAGGCTGAA	780
	CAGATAAAATC AGGCAGCAGG AGAGGCCAGT GCAGTTCTGG CGAAGGCCAA GGCTAAAGCT	840
	GAAGCTATTG GAATCCTGGC TGCAGCTCTG ACACAACATA ATGGAGATGC AGCAGCTTCA	900
15	CTGACTGTGG CCGAGCAGTA TGTCAGCGCG TTCTCCAAAC TGGCCAAGGA CTCCAACACT	960
	ATCCTACTGC CCTCCAACCC TGGCGATGTC ACCAGCATGG TGGCTCAGGC CATGGGTGTA	1020
20	TATGGAGCCC TCACCAAAGC CCCAGTGCCA GGGACTCCAG ACTCAGCTCTC CAGTGGGAGC	1080
	AGCAGAGATG TCCAGGGTAC AGATGCAAGT CTTGATGAGG AACTTGATCG AGTCAAGATG	1140
	AGTTAGTCCA GCTGGGCTTG GCCAGGGAGT CTGGGGACAA GGAAGCAGAT TTTCTGATT	1200
25	CTGGCTCTAG CTTCCCTGCC AAGATTTGG TTTTTATTCTT TTTATTTGAA CTTTAGTCGT	1260
	GTAATAAACT CACCAAGTGGC AAACCAAAAA AAAAAAAA AAAAAAAA AAAAAAAA	1320
	AAAAAAA AAAANNN	1337
30		

35 (2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1276 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

45	CTCACGGTCA CGGGACGGCN GGACGCGTGG GTGCATTTC TGAGTGTCTT ACTTCCAATT	60
	ATGTGATTCTN ATATTACAGG NGCTGCCATG TGGTAATGAG AAGAATGTAT ATTCTGTTGT	120
	TTTGGGTTGG ARTGTTCCAT AGATGTCTAT CARGTCTGTT TGATCCAGAR CTGARTTCAR	180
50	GTCTGGTAT CTCARTCTTT ACTGTGARTC TTCAAATGAC ATAAGAATGA CAGAAMTTGT	240
	AGTTAAGGAC AACAGRGCAW TSCAAGGCAG CAGCATAGTC CAAAATAGAC GTGTCTCTT	300
55	CCCGAAGTCA CTGTAGTGGG GGACATAAAA TTAAAGGAAC CTCTGGGTCT TACTACCTGA	360
	TGTGGCCAAT TGGACTAAAA CCAATAACCA TTAAGGAAWA AATSSACTWA ACCACAAGCA	420
	ACTCAATTAA MAATAGGCA AAGAACATGA AGAGGCATT TCCCAAAGAA GCCAACAAAGC	480
60	ATGTGAAAAG ATGCTCAACA TCATTAGACA TCAGGGAAAT ACAGATCAAATCAAAATGA	540

	GATACCAGTT TATACTAAGG TGGCTATAAT AAACATCATA ATAATGAAGG ACATTAACAT	600
5	GTATTAGTGA GGATGTGGAG AAATGGAACC CATTCTGGT AGGAATGTAAT AATAGTGCAG	660
	CCACTGTGGA AAACAGTTG GTGGTCCCC AGAAAGCTAA GCATAGAGTT ACCAGAGAAC	720
	CTAGCAATTAACTTATAGG TACATACATTC AAAGGAATTG AAAACATAGA TYCTAACAGA	780
10	TACTRGTACA GCAATATYCA TKGTGGCWT ATTACGATA GCCAAAAGGT AAAACAAC	840
	AAGTGTCCAT CAAAATATAA ATGTGTAAAC AATGTGGTAT ATTCTAGAG GGGAAATATTA	900
15	TTCAGCTTTA AAAAGGAATG AAGTACTGGT ACATGCTACA AAGGTGGATG AGCCTCAGAA	960
	ACATGCTGAG TGAAAAGAAC CAATGATAAA AGACCATATA TTGTATGATT CCATTATATG	1020
	AAATKTCCAG RACATTCAAG TCTATAGAGA CAGAAAGTAG ATTAGTGAAT GCTTAGGGCT	1080
20	GGCAGGGATA AGGGGKTCAT GGCTAAAGGG TATGGGTTTG TGTTTGTGGA GGTGAAAAAT	1140
	TTTAAAACCTT GKGSSTGATGG TTGACAAGC CTGTGAAGAT ACTGAAAACC ATTGAATTGT	1200
25	GTGCTTTAAA TCGATGAATT GTATGGTGTGTT TGAACATATAT CCCAATAAAAG CTGTTTTTA	1260
	AAAAAGAAAA AAAAAAA	1276

30

(2) INFORMATION FOR SEQ ID NO: 47:

	(i) SEQUENCE CHARACTERISTICS:	
35	(A) LENGTH: 1282 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
40	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:	
	GGCACGAGAG AAAGGCCAGT TTGTGGGCA AATTAGACTA AACCTCTGTGC TGGTAGAACT	60
	GCTTTCCAAG AATGCTGTCA CTGCTATAGT TTTTAATGCT TCAAATCTCA ACTCNCCTCC	120
45	TCCATTGCCCC ATAGCTAACAC CATGTTCCAG GAGTGTATTCA CAATCAGCTT GTTTTTCTTT	180
	AACTGGTCAA AGGAATGTTG CTCATTCAAC TGCCCCAACT CACATATTTAA CAATTGTTTA	240
	ACTGGGATTA GATAAAAGGA AAGCTGACTT ACAGATGAAC CAAGAGGGAG CTATTATG	300
50	CACAGCCCCC AGCCCAGTAA CTTTATGTTT CTGATCTCCT GCAAAATTTT TTTATAAAAAA	360
	AAGCTTAGCC AGGAACCTAGT AGAAAGAATA AAGTAAAGAT GGTGTAAGAA ATATATGGAT	420
55	AGGCAAGTTC CWNYGYTGAG ACCTTAYGAA GAATGGTGAG GTGTGGTAA ATGGAGGAGA	480
	TAATCAGCAG ATAAWAGCTC AGATGGTCMS AAACATWTAG AACTATAATG CCATCTCCAA	540
60	AGTATTGCAT GCATACAAAT GACGTTCAAT CCGTTGAATA TAATGGAGAC ACACATTTTC	600

	AAAAAATTAAAG TTCTTCTWTC TTGAGCTTTA AAAGTATACA CATTACCCM AATGAATTWA	660
	AAACATGCMC ACMATATTTT ATATCAAAG TGTACATGAT TTCCAAAAGT TGGAAGTWAC	720
5	CAAGATTTAC TTCCWTGGGT TAGTGCATAA ATTAACGTG ATACATATAT ACTATGGAAT	780
	WTTAYTCAGC AACAGAAATA AATGAGHTAT CAAACCACAG AAAGACATCG AGGAAACTTA	840
10	AATCCAGGTG GMTAAGTGAW AGAAGCCAAT ATGAAAAGGC TACATTSTAT ATGATTTCAA	900
	ATATATGACA TTICAGGAAA GGCAAGGCTG CAGAGACAGT AAARAGATCA GCTAGGTGCA	960
	TGKGGSCTCAC GCCACTTTGG GAGGCTTGAG GCAGGGGAT TATMTGAAG TCAGGAGTTTC	1020
15	NAGACCAGCN TGGCAACAT GNTGANACCC CATATNTCCT AAAAGNACNA AAATTAACT	1080
	GGGGCTGGTG GCACGTGCCT GTANTCCCAN CNACTCTGGT GGCTNAGACN GGNGAATTGC	1140
20	TTGAACCCAG GAGGCAGAGG TTGCGGTGAG CCAATGATTG CACCACTGCA NTCCAGCCTG	1200
	GGTGGTAGAG CGAGACTCAG TCTCAACNTT NATCAAGATA GGANNGAAT AGAANGGAAG	1260
	AAAGAGAAAA AATAAAAATA NA	1282
25		

(2) INFORMATION FOR SEQ ID NO: 48:

30	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 645 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:	
	AAGGTAGAAA AGTACAGAAA ACACAAATT TTCATTGTGC TGTTCATG TGCCAGATTC	60
40	TTTAAATAC TTGACACCC TACAATAATT AAAGGTTTA AGAACATTA GATACTTTAA	120
	AAATAAAAGC CCACAATTGA ATAACAAAAA TGAACCTTGT TTTATTTTTT ATTGGCATTA	180
45	ATGTAGGTTG CCGTGGTGAA AATAGTTGA AATACCTCAC AGTAACAGTT TTGTGCAGCC	240
	CTAGAGATTA AAAACAGCAA AGTAAATAAG CAGGACTCTC AACGACTCAT ACTCACAGAC	300
	ATGTTTAATG TAATCCTAGC ACTTCGGGAG GCTGAGGCGG GAGGATTACT TGAGCCTAGG	360
50	AGTTTGAGAC CAGCCTGGC AACATAGCAA GATCCCCTCT CTACAAAAAA GTGAAAAAGT	420
	TAGCTGAACA AGGGGGCATG CACATGCTAC TCCAGACGCT GAAGTGGGAA GATCCTTAA	480
55	GTCCGAGAGA TCGAGGCTTC AGTGAGATAT GGCTGAGACA CTGCTCTAG CCTGGATGAC	540
	AGAGTGAGAA CCTGTCTCAA ACAAGAGAAA AAAATAAATC AAATGCTATT CAAATCTA	600
	AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAA	645
60		

(2) INFORMATION FOR SEQ ID NO: 49:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1495 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TGTGGAAAAC	AGTAGGAAAG	CAATGAAAGA	AGCTGGTAAG	GGAGGCCGCG	CTGATTCCAG	60	
15	AGAGCTAAAG	CCGATGGTAG	GTGGAGATGA	GGAGGTGGCC	GCCCTCCAAG	AATTCACTT	120
	TCACTTCCTC	TCTCTCTCTG	TCTTCACTGA	CTGCACTTCT	TCAGGAGAAG	CTTTTGTAT	180
20	CTGTATCACG	CAGACATGCT	GCTCTTCTG	TTTGTGTGCT	TACCCATCAC	TTGGATGGCA	240
	GAATTCTTGT	CACAACGTGAG	ACCACCTCT	ATAAAAGTAA	CCTGAAAGGA	ACACCATCCT	300
	CGTCAGTGCT	CGGCAGGGGC	GGGTAGGGGA	TGATGGTTTT	TTCCCTAAGG	TAAAACGTCT	360
25	GTTGCTCTTG	TTTCTTTTTT	AACTGTCAGT	GTTTGGCTTT	CATCAGAMTG	AACATTTGG	420
	TGTTCCACTT	GAACTGACGG	TTTGATTTTT	ATCATTGTTG	AAAGGTGATC	ATAGCAATTG	480
30	CTTTCCAATC	TGCTAAAATT	CCATACTCCC	CCCTTTAA	ARWATKGTS	TGCTTMCATT	540
	GCTKIMCWIT	TSCCTTGKCT	SMCTTTTTCY	TCCCTGKGSC	TGAARTTKTW	CYTTCYTTKT	600
	TTCTTAAGST	WTTTTTCAGT	AGCAAACAAG	GCTGTTTCA	TCAATACCCA	CATTCCAYT	660
35	CRGKRRGRMM	ATYTAGTYTT	YTCCCAGKTT	AAKIGKGRGR	KGGRKGAAAA	TRATKTCGG	720
	KANGKGGAWA	TKAWAWAKGR	KWWATGKAAA	CACAAATATA	TYTYTYTAMA	TTCCACTTTA	780
40	ATTKGAAA	AAAGGCAGCT	KAAGTGGAGT	GTWAAGRARR	ACCTKGRRST	GCTTTCAAC	840
	ATGGGATATG	GTCACTATRG	CATRGAAAC	ANGATGCCCT	CTATCAWAKA	TGGGTCTAAT	900
	TACTYCCTAA	TTTAAACAC	GTATTTTTTT	AAATAGCATG	TTTATTTTCA	AATATDATAT	960
45	AATGGTCGSG	CRCCCTTAAA	TAATTTAAA	CAANGTGCC	CCGRGACNGC	ATATAATGTT	1020
	CAAANGTKAG	AGGTAAGGAC	TTYCCPTCT	GTCTYCTTAA	CACTIWAGTA	AATRATTNGA	1080
50	WTTAWACCAA	GTGGTCCAA	CTKGNNCCCT	GNGGNCCCA	NANGGMGRG	GAAGGGCTTT	1140
	TCMAACACAA	ATTCTGAAAC	TTTATTTAAA	CATGAGATTT	TTTGCCTTTT	TTTTTTTAAG	1200
	CCCATCAGCT	ATCCTTAATG	TATTTTANAT	GTGGCCCAAG	ACAATTCTTC	TTCCAGGATG	1260
55	GCCTGGGAA	GCCAAAAGAT	TGGANACCCC	TGATTTGTAG	GTTTCAACT	TTAAAATATA	1320
	TGCTATAAAA	TAAGTTCATT	TAAGTAGGCT	AGGCATGGTG	GTCATGTNT	GTAATCCTAG	1380
60	CACTTACGGG	GCCCGAGGCA	GAAAGATTRM	CTGAGCTCAG	CAGTTGAGA	CCAGCCTGGG	1440

CCAAACGGTG	NAACCCGTGTT	TTTACTNAAA	TACCCAAAAA	AAAAAAAAAA	AAAAAA	1495
------------	-------------	------------	------------	------------	--------	------

5

(2) INFORMATION FOR SEQ ID NO: 50:

	(i) SEQUENCE CHARACTERISTICS:	
10	(A) LENGTH: 1630 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:	
	GAATTCGGCA CGAGATTATC TGTCTTCCTTC TTACCAATTG ATAGAACCTTT TTAGTATTGC	60
	AGATAAAGTT CCTCATCGGA TATCTTCTCT CCTTCTATTG GGTACCTTTT TATTGCTTA	120
20	ATGGGGCTCT TTTAATGACC AGAAGTCCTT AGTTTAAAAA TAGTCCAGTT TATCCATTTC	180
	TAAATTGTTA GTGCTATTG TGTCTGCTT GAGAGATTTC TGCTACTGC AAGGTACAA	240
25	AGATGTTTTC CTCTAAAAGC CTTTGGTTT TGCCCTTTG TTTAGATCT GCACCTCATC	300
	TGGAATTGAG TGTGTGGTGT GTGTGTGGTG TGAGGTAGGG GTCCCTTTTT TCATATGGAT	360
	ATCCAATTGAA CCCAGAACAG TGTATTGAAA AAAAAAATCT GTCTTAGTCA ATTTGGACTG	420
30	CCGTAACAAA ATACCATAAC CTGGGTGGCT TAGACTACAG AAATGTAGCG CTCACAGYTC	480
	TGGAGGCTGG AAGGCCAGGA TCAAGACACC AGCAGATTG TGTCCTNGTG AGGACCCACT	540
35	TTGTGNITCA TAGATGTCAC CTTCTTGCTG TGCCCCAGTG GTGRAAGGGG CAAACTAGCT	600
	CCCTTAAACC TCTTTTATA AGATCCCTAA AACCTTTAAT GAGGGCTCCA CCCTAATGAT	660
	CTAACACCT CTCAATACCT TATCTTGGGG GTTAAGATTT GAACAGAGGA ATTTGGGGGA	720
40	GACATAGACA TTTGGAGCAT AGCATCTCTT TTTCTCAGT GCACAGCAGT GCTGCCCTCA	780
	TCATCAGTCA GGTGTCTGTA GGTGTGTGGC TATTTCTGGA CTTGGCACTC TGTCCTACTT	840
45	GTTGATTCTCT CGCCTTATA CCAATGCCAC ACCATCTTAA TTATTGTAAC CATCTTAATT	900
	ATTTATAAAA AGTCTTTTTT TTTTTTTGTA TACAGTCTCA CTCTGTCCCC CAGGCTGGAG	960
	TGCAGAGGTA CACTATTGGC TCACTGCAAC CTCTGTCCCC AGGCTTAAGC AATTCTCATG	1020
50	CCTCAGCCTC CTGAGTAGCT GGGATTACAT GTGCACCACC ACACCTGGCC TTCTTCTTT	1080
	TCTTCTCAAY CCATTKGTTT TTTATTCCTT TCCCTKGCTT TATKGCACTG GCTAAGATTT	1140
55	CCAGTGTGA ATAGGAGTGA TGACAGTGGG CACCCCTGTC TTCTCTCCAA CCTCAGAGGG	1200
	AAAAGTATCC AATGCATTG TAGATATTCT TTATCAGATT AGCTTCCCTT CTAGCGGCTT	1260
	GIGTCCTTGC ATGTTTTTC ATGAGCAAGT GTTGAACCTT TTCACTGAGT TTTCAAATA	1320
60	CTTTTCCAT TGAGTTTTT TACTTTAACG GTCAATTGCA CAAAAGTCTG CATTGTTAT	1380

	TTCCCTCCCAA ATTGCTGGGA TTATAGGCAT TAGCCACTGC ACCCAGCCAG ACTTTATAGA	1440
5	AAATCTTGTAT ATCTGGTCAT GGAAGTCCCC TAGCTTGGTT ATTTTTTTT GGTACCGCTT	1500
	TGTCTATTTT CGGCCCTTTC CATTTCATG TAACTTTAG GATCAGCTTG TCAGTTCTA	1560
	CCAAAAAAA AAAAAAAA ACTCGAGGGG GGCCCCGTAC CCAAATGCC GGGTAGTGAT	1620
10	CCTAACAAATC	1630

15 (2) INFORMATION FOR SEQ ID NO: 51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2420 base pairs
- (B) TYPE: nucleic acid
- 20 (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

25	GCCAACAGTG CTCCCTCATA GATGGACGAA GTGTGACCCC CCTTCAGGCT TCAGGGGAC	60
	TGGTCCCTCCT GGAGGGAGAT GCTCGCCTTG GGGAAATAATC ACTTTATTTG TTTTGTGAAT	120
30	GATTCTGTGA CTAAGTCTAT TGTGGCTTTG CGCTTAACTC TGGTGGTGAA GGTCAGCACG	180
	WGGCCGGGGG AGAGTCACGC AAATGACTTG GAGTGTTCAG GAAAAGGAAA ATGCCACCACG	240
	AAGCCGTCAAG AGGCAACTTT TTCCCTGTACC TGTGAGGAGC AGTACGTGGG TACTTCTGT	300
35	GAAGAAATCG ATGCTTGCCA GAGGAAACCT TGCCAAAACA ACGCGAGCTG TATTGATGCA	360
	AATGAAAAGC AAGATGGGAG CAATTTCAACC TGTGTTGCC TTCCCTGGTTA TACTGGAGAG	420
40	CTTTCGCAGT CCAAGATTGA TTACTGCATC CTAGACCCAT GCAGAAATGG AGCAACATGC	480
	ATTTCCAGTC TCAGTGGATT CACCTGCCAG TGTCCAGAAG GATACTTCGG ATCTGCTTGT	540
	GAAGAAAAGG TGGACCCCTG CGCCTCGTCT CCGTGCCAGA ACAACGGCAC CTGCTATGTG	600
45	GACGGGGTAC ACTTTACCTG CAACTGCAGC CGGGGCTTCA CAGGGCCGAC CTGTCCCCAG	660
	CTTATTCGACT TCTGTGCCCT CAGCCCCCTGT GCTCATGGCA CGTGCCGCAG CGTGGGCACC	720
50	AGCTACAAAT GCCTCTGTGA TCCAGGTTAC CATGGCCTCT ACTGTGAGGA GGAATATAAT	780
	GAGTGCCTCT CGCGTCCATG CCTGAATGCA GCCACCTGCA CGGACCTCGT TAATGGCTAT	840
	GAGTGTGTGT GCCTGGCAGA ATACAAAGGA ACACACTGTG AATTGTACAA GGATCCCTGC	900
55	GCTAACGTCA GCTGTCTGAA CGGAGCCACC TGTGACAGCG ACGGCCTGAA TGGCACGTGC	960
	ATCTGTGCAC CGGGGTTTAC AGGTGAAGAG TGCGACATTG ACATAAAATGA ATGTGACAGT	1020
60	AACCCCTGCC ACCATGGTGG GAGCTGCCTG GACCAGCCCA ATGGTTATAA CTSCCACTGC	1080

	CCGCATGGTT GGGTGGGAGC AACTGTGAG ATCCACCTCC AATGGAAGTC CGGGCACATG	1140
	GGGGAGAGCC TCACCAACAT GCCACGGCAC TCCCTCTACA TCATCATCTGG AGCCCTCTGC	1200
5	GTGGCCTCA TCCCTATGCT GATCATCCTG ATCGTGGGA TTGCGGCAT CAGCCGCATT	1260
	GAATACCAAGG GTTCTTCCAG GCCAGCTAT RAGGAGTTCT ACAACTGCCG CAGCATCGAC	1320
10	AGCGAGTTCA GCAATGCCAT TGCATCCATC CGGCATGCCA GGTTTGGAAA GAAATCCGG	1380
	CCTGCAATGT ATGATGTGAG CCCCATCGCC TATGAAGATT ACAGTCTGA TGACAAACCC	1440
	TTGGTCACAC TGATTAACAC TAAAGATTTC TAATCTTTT TTGGATTATT TTTCAAAAG	1500
15	ATGAGATACT ACACATCTT AAATTTTT AAGAAWAA AAAGCTTAAG AAATTTAAAA	1560
	TGCTAGCTGC TCAAGAGTTT TCAGTGAAT ATTTAAGAAC TAATTTCTG CAGCTTTAG	1620
20	TTTGGAAAAA ATATTTAAA AACAAAATTG GTGNAACCTA TAGACGATGT TTTAATGTAC	1680
	CTTCAGCTCT CTAACATGTG TGCTTCTACT AGTGTGTGCT CTTTCACTG TAGACACTAT	1740
	CACGAGACCC AGATTAATTCT GTGTTGTGT TACAGAATAA GTCTAATCAA GGAGAAGTT	1800
25	CTGTTGACG TTGAGTGCC GGCTTCTGA GTAGAGTTAG GAAAACCACG TAACGTAGCA	1860
	TATGATGTAT AATAGAGTAT ACCCGTTACT TAAAAAGAAG TCTGAAATGT TCGTTTGTG	1920
30	GAAAAGAAC TAGTTAAATT TACTATTCTT AACCCGAATG AAATTAGCCT TTGCTTATT	1980
	CTGTGCATGG GTAAGTAAC TATTTCTGCA CTGTTTGTGTT GAACTTGTG GAAACATTCT	2040
	TTCGAGTTTG TTTTGTCTT TTTCGTAACA GTCGTCGAAC TAGGCCTCAA AAACATACGT	2100
35	AACGAAAAGG CCTAGCGAGG CAAATTCTGA TTGATTGAA TCTATTTT TCTTTAAAAA	2160
	GTCAAGGGTT CTATATTGTR AGTAAATTAA ATTTACATTT GAGTTTTTG TTGCTAAGAG	2220
40	GTAGTAAATG TAAGAGAGTA CTGGTTCCCTT CAGTAGTGAG TATTTCTCAT AGTGCAGCTT	2280
	TATTTATCTC CAGGATGTTT TIGIGGCCTG ATTGATGTA TATGTGCTTC TTCTGATTCT	2340
	TGCTAATTTC CAACCATATT GAATAATGT GATCAAGTC AAAAAAAA AAAAAAAA	2400
45	AACTCGAGGG GGGGTCCCGT	2420

50 (2) INFORMATION FOR SEQ ID NO: 52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1172 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

60 AAAATTATTC TGTACCATCA CAGCTTTCA CAACGATGGC AAGCCTTATG TCTTGGGAGC

60

	CTGTTTGCT AGGCAAAGTT ACAAGTGACC TAATGGGAGC TCAAATGTGT GTGTGTCCT	120
5	CTGTGTGTTT GTGTGTGTGT GTGCACTCAA GACCTCTAAC AGCCTCGAAG CCTGGGGTGG	180
	CATCCCGGCC TTGCCATTAG CATGCCATCAT GCATCATCAG ATGACAAGGA CAACCTCAT	240
	GACGAAGCAA CATGAATTAG GGGCCTCTT GGCCCTGGTC CAAAATTGTC AATCAGAAAT	300
10	GAACATAAAG GACTCCAGAG CAGTGGACT GTCTGTCAAAGACTCTGTA TATCTTTGT	360
	GGATGAGTTT TGTGAGAGAA CAGAGAGACC ATTGTACCTG GCACAAGGGC TSTTCATGAA	420
	AACGGGAGCT TACTGGGAGG TCCAAGACAG TGGCATTCT CCTCTCCCT TGCTGCTCAG	480
15	CACAGCCCTG GATTGAGGCC CCGAGGCTGA GACCAGACAA AGCCCGGGAG GCAGAAAGAT	540
	GCTCCAAGAA CCAACACTAT CAATGTCTTT GCAAATCCCTC ACAGGATTCC TGTGGGTCCA	600
20	GCTTGGAAC TGGGAAACCT TTCTTCGGAT CGCGACTCAT TCCACTGATG CCAGCTGCC	660
	CTGAAGGATG CCAGTACTGT GGTGTGTGAG TCTCAGCAGC CGCCCACACG CTCTAACTC	720
	TGCTGCAATGG CAGATGCCATA GGTGGAAATA GCAAAACAA GCCCCAGGCT GGGGCCAGGG	780
25	CCAGAGGGGA AGGCCCTGGA TTCTCACTCA TGTGAGATCT TGAATCTCTT TCTTTGTTCT	840
	GTTTGTTTAG TTAGTATCAT CTGGTAAAT AGTTAAAAAA CAACAAAAAA CTCTGTATCT	900
30	GTTTCTAGCA TGTGCTGCAT TGACTCTATT AATCACATT CAAATTCAAC CTACATTCC	960
	CTCCCTCTCA CTAGCCTCTC TGAAGGTGTC CTGGCCAGCC CTGGAGAACG ACTGGTGTCT	1020
	GCAGCACCCCC TCAGTTCTG TGCCTCAGCC CACAGGCCAC TGTGATAATG GTCTGTAG	1080
35	CACTCTGTA TTATGTAA GAATGATTAT AATGAAGATA CACACTRTAA CTACAAGAAA	1140
	TTATAAAATGT TTTCACATC AAAAAAAA AA	1172
40		

(2) INFORMATION FOR SEQ ID NO: 53:

45	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1589 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:	
	CCACCGCGTC CGCCCCACGCG TCCGCCACG CGTCCGTTTC AAAGGGAGCG CACTCCGCT	60
55	GCCCTTTCTT TCGCCAGCCT TACGGGCCCC AACCCCTCGTG TGAAGGGTGC AGTACCTAAG	120
	CGGGAGCGGG GTAGAGGCGG GCCGGCACCC CCTCTGACC TCCAGTCCCG CGGGCCCTCAA	180
60	GATCAGACAT GGCCAGAAC TTGAAGGACT TGGCGGGACG GCTGCCCGCC GGGCCCCGGG	240

	GCATGGGCAC	GGCCCTGAAG	CTGTTGCTGG	GGGCCCCCGC	CGTGGCCTAC	GGTGTGGCG	300
	AATCTGTGTT	CACCGTGGAA	GGCGGGCACA	GAGCCATCTT	CTTCAATCGG	ATCGGTGGAG	360
5	TGCAGCAGGA	CACTATCCTG	GCCGAGGGCC	TTCACTTCAG	GATCCCTTGG	TTCCAGTACC	420
	CCATTATCTA	TGACATTCCG	GCCAGACCTC	AAAAAATCTC	CTCCCCCTACA	GGCTCCAAAG	480
10	ACCTACAGAT	GGTGAATATC	TCCCTGCGAG	TGTTGTCTCG	ACCCAATGCT	CAGGAGCTTC	540
	CTAGCATGTA	CCAGGCCCTA	GGGCTGGACT	ACGAGGAACG	AGTGTGCG	TCCATTGTCA	600
	ACGAGGTGCT	CAAGAGTGTG	GTGGCCAAGT	TCAATGCCCTC	ACAGCTGATC	ACCCAGCGGG	660
15	CCCAGGTATC	CCTGTGTGATC	CGCCGGGAGC	TGACAGAGAG	GGCCAAGGAC	TTCAGCCTCA	720
	TCCTGGATGA	TGTGGCCATC	ACAGAGCTGA	GCTTTAGCCG	AGAGTACACA	GCTGCTGTAG	780
	AAGCCAAACA	AGTGGCCCAG	CAGGAGGCC	AGCGGGCCMA	ATTCTTGGTA	AAAAAGCAA	840
20	ACCAAGAAC	GGGGCAGAAA	ATTGTGCAGG	CCGAGGGTGA	GGCCGAGGCT	GCCAAGATGC	900
	TTGGAGAAC	ACTGAGCAAG	AACCTGGCT	ACATCAAAC	TCGCAAGATT	CGAGCAGCCC	960
25	AGAATATCTC	CAAGACGATC	GCCACATCAC	AGAACCGTAT	CTATCTCACA	GCTGACAACC	1020
	TTGTGCTGAA	CCTACAGGAT	GAAAGTTCA	CCAGGGGAAG	TGACAGCCTC	ATCAAGGGTA	1080
30	AGAAATGAGC	CTAGTCACCA	AGAACTCCAC	CCCCAGAGGA	AGTGGATCTG	CTTCCTCCAGT	1140
	TTTTGAGGAG	CCAGCCAGGG	GTCCAGCACA	GCCCTACCCC	GCCCCAGTAT	CATGGATGG	1200
	TCCCCCACAC	CGGTTCCCTG	AACCCCTCTT	GGATTAAGGA	AGACTGAAGA	CTAGCCCCTT	1260
35	TTCTGGGAA	TTACTTTCT	CTTCCCTGTG	TTAACTGGGG	CTGTTGGGA	CAGTGGTGA	1320
	TTTCCTCACTG	ATTTCTACA	GTGTTGTTCC	CTCCCTCAAG	GCTGGGAGGA	GATAAACACC	1380
40	AACCCAGGAA	TTCTCAATAA	ATTTTTATTA	CTTAACTGTA	AGTCAAGGCT	TCACGTGTT	1440
	ATGAACTGGG	TAACTGGCAG	CAAGCATGGC	CACGGTCA	TGTGGCTCC	TGGGTCTGTC	1500
	TTTGTGTG	CCACCAAGGGG	GCGCAAAAGA	ATCTGGCTGG	GGCGGCTAAN	GGGAAGCAAG	1560
45	GCCTGGCTC	CGAAACANGA	CCCAACTGG				1589

50 (2) INFORMATION FOR SEQ ID NO: 54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2074 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

60 CGGCCTGACC

60

GCCCCGGGCT

TAAGGGAGCC

TGGCTAGGCC

GGCAGCCCGA

TGGTCCCGCA

	GCTCGGGGCC	GGCCATGCCT	CCCGGTCCGT	GGGCCAGCT	TTGGCTCTTT	YTCCCTGCTGC	120
5	TCCTCCCGG	CGGGCCTGAG	CCCGCGGGG	CCTCCAGGCC	GTGGGAGGGA	ACCGACGAGC	180
	CGGGCTCGG	CTGGGCTGG	CGGGCTPPCC	AGGGCCTGCA	GGAGCAGCTC	AGGGCGGGG	240
	GTGCCCTCTC	CAAGCGGTAC	TGGACGCTCT	TCAGCTGCCA	GGTGTGCC	GACGACTGTG	300
10	ACGAGGACGA	GGARGCAGCC	ACGGGGCCCC	TGGCTGGCG	CCTTCCTCTG	TGGGCCAGC	360
	GGTACCTGGA	CCTCCTGACC	ACGTGGTACT	GCAGCTCAA	AGACTGCTC	CCTAGAGGGG	420
15	ATTGCAGAAT	CTCCAACAAC	TTTACAGGCT	TAGAGTGGGA	CCTGAATGTG	CGGCTGCATG	480
	GCCAGCAATT	GGTCCAGCAG	CTGGTCCTAA	GAACAGTGAG	GGGCTACTTA	GAGACGCC	540
	AGCCAGAAAA	GGCCCTTGCT	CTGTCGTCC	ACGGCTGGTC	TGGCACAGGC	AAGAACCTCG	600
20	TGGCACGGAT	GCTGGTGGAG	AACCTGTATC	GGGACGGGCT	GATGACTGAC	TGTGTCAGGA	660
	TGTTCATCGC	CACGGTCCAC	TTTCCCTACC	CCAAATATGT	GGACCTGTAC	AAGGAGCAGC	720
25	TGATGAGCCA	GATCGGGAG	ACGCAGCAGC	TCTGCCACCA	GACCCCTGTT	ATCTTCGATG	780
	AAGCGGAGAA	CCTGCACCCA	GGGCTGCTGG	AGGTCCCTGG	GCCACACTTA	GAACGCC	840
	CCCCCTGANGG	CCACAGGGCT	GAGTCTCCAT	GGACTATCTT	TCTGTTTCTC	AGTAATCTCA	900
30	GGGGCGATAT	AATCAATGAG	GTTGGCCTAA	AGTTGCTCAA	GGCTGGATGG	TCCCGGGAAG	960
	AAATTACGAT	GGAACACCTG	GAGCCCCACC	TCCAGGCGGA	GATTGTGGAG	ACCATAGACA	1020
35	ATGGCTTTGG	CCACAGCCGT	CTTGTGAAGG	AAAACCTGAT	TGACTACTTC	ATCCCCTTCC	1080
	TGCCTTTGGA	GTACCGTCAC	GTGAGGCTGT	GTGCACGGGA	TGCTTCC	AGCCAGGAGC	1140
	TCCTGTATAA	AGAAGAGACA	CTGGATGAAA	TAGCCCAGAT	GATGGTGTAT	GTCCCCAAGG	1200
40	AGGAACAAC	CTTTTCTTCC	CAGGGCTGCA	AGTCTATTTTC	CCAGAGGATT	AACTACTTCC	1260
	TGTCATGAAG	GCTAGAGGAA	GACTTCCCTGG	AACTGCCTTT	CTTCCACTAA	CAGGACCC	1320
45	GGACCTGTAG	GAGCACCCCG	TTTGGGACTG	TGAGGTGT	GAGGGTGTGG	ACTGGCATCC	1380
	AGCAGCCACT	AACAAACACA	CAACTGGGT	GTAAAAGGCA	GGCCTTACAT	TAGAACCAA	1440
	GCCAATCCTT	TTTCTTTTTT	TGGAGGTCC	CACCGAGATA	GATAGGAAC	TGGATTGCTG	1500
50	AAATTCAAAA	CAGAGCCAT	TCTTAAGATC	ACTTGGTGC	TTAAAGACAC	GCATTCAA	1560
	GTGGAATGTG	GTGGAAGAAA	GTGGGCCAGG	TGGTTGAAGA	AAGCCATGTG	GGAGCTCAGC	1620
55	AAATCCCAAG	GGCTTATTAT	GACACTCCAG	ATGGCTCCT	TAGCATCTCA	GCTCTTCTGC	1680
	AAGGAAGAGC	TTGGGTGTTA	GGCCTCAGAG	GCTGTAGGGT	CCTTGGGTTA	CAGAGCCGGG	1740
	GAGAACGAAG	TTCGTGACC	CAGGGGTGGA	GAATACACTC	TAGGTTGCG	GGCTGGTGG	1800
60	CTTCAAATT	GGTACTTCCA	GAGGAAAGCC	AAGCTGCTTC	TGTTGTGAGC	GAATCAGCCA	1860

	AGAGCCTGAG GCTGAAGGGA AAAGTACACA GAGGAAGATA TTTTACAAAC CAGGTAGTG	1920
5	TAGGCCAAGA CTTATGGTCT ACAGATTTTG CGGGGGGAGG GGGGACCTTT TCAAAGACAA	1980
	TAGGGGTCT TGACATGTTT GTTGTATGTA AAGATGATAA GATTAATT TTTGATTTTC	2040
	CTAAAAAAA AAAAAAAA AAAAAAAA TTNC	2074

10

(2) INFORMATION FOR SEQ ID NO: 55:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1483 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

	GAATTTCGSCA CGMCGGTGGA GGCGCCACGT CCCCTGCGGC GCGGGGAGAG AAATCGCTTG	60
25	GACTTCGGGG CGGCCTCGGA CGGCCATGGC CTTTACCCCTG TACTCACTGC TGCAGGCASC	120
	CCTGCTCTGC GTCAACGCCA TCGCAGTGCT GCACGAGGAG CGATTCCCTCA AGAACATTGG	180
30	CTGGGAAACA GACCAGGGAA TTGGTGGATT TGGAGAAGAG CCGGAAITTA AATCACAGCT	240
	AATGAACCTT ATTCGATCTG TAAGAACCGT GATGAGAGTG CCATTGATAA TAGTAAACTC	300
	AATTGCAATT GTGTTACTTT TATTATTTGG ATGAATATCA GTGGAGAAAA TGGAGACTCA	360
35	GAAGAGGACA TGCCAGTAGA AGTTATTACT TTGGTCATTA TTGGAAATATT TATATCTTAG	420
	CTGGCTGACC TTGCACTTGT CAAAAAATGTA AAGCTGAAAA TAAAACCAGG GTTTCTATTT	480
40	ATCTGTTTTT TTTTTTAATG TTGCACTTGT AGTTTCATTA CAAAAGATCA GATCATGAAA	540
	GGCAGTAACT CTCCAGGACT GGAATATCTG ATTGCTCAGT GTTAATAGTA GTTCATGCTG	600
	TGGTGAGATT GTAAAAGGG TGCAAGACTG TTGCTTCTCT TTTTTAGAT ATTTTCTAT	660
45	CTCTCACCTC TCAGGGATGA AATTCTTTTG CAAAGTTTG AAGTTCTTG CAACTAGCC	720
	ATGATGTGAG TGGTTATCCC TAGATAAAAT TAAAAGGATT TTTAAAAGT AATTACTGCA	780
50	CATAAAATGA TAAATAGGTA ATTTGAATAA TTTTATTTA AGCTCCTTG TTAATTATTT	840
	TGTCTATTGT CTCAGCTATA AATTCAAATT TATACATACT ATTGAGTATT AATATTCTCT	900
	GATTTCAGGG AGAATTCTGT CAGTCACATG ATGATTATGT TTTTNTTTAA CATTCTTTCC	960
55	ATGCACTTGT TATTTTATTA ATTTGCCTGA ATGATGAGAC CAGACCAGTG TCTACAGATT	1020
	TTCATGTCA GAAAATCTA TAAGCTGCC CTTTTACAA TGATGGATT AAAAACA	1080
60	ACAGCGTAAA TATTAGCCCA CAAGAGCAGT CCTAAACAAT CACAATTACA CTGTACTACC	1140

	CAAGAAGACT GTTTATTGTG AAGCATTAC CTTTCAAAAA ATCATTACAT TTCTATTCT	1200
	TGGTGGAGCA GCACATTGTG GAGTGTGATT CTTAATTCTT CATTGAGTTT GTCAATAGGA	1260
5	CATTGATGCT GGATAGGTG TCTTTGTTT TTATGTCTCA GACCATCTG TGAGATTGTT	1320
	TGCCATCTC ATAATACAGT TTTATGCAGA AAGGTTGAAA CTATGTAAT GGTTTTATG	1380
10	GAAATTATCA GTTACAATAT TTTAAAGGTG TAGAATGGCA TCTTGTATA TAGGAGAAC	1440
	TTTGTAAATA AAGTTAAATT TCTAAGTCAA AAAAAAAAAA AAA	1483

15

(2) INFORMATION FOR SEQ ID NO: 56:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	1123 base pairs
(B) TYPE:	nucleic acid
(C) STRANDEDNESS:	double
(D) TOPOLOGY:	linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

25	CAAAAATAAT AATAGTCATC ACAATTGTAT AGCACTGGGT CATTTCCTCC AAGACCAATT	60
	AGTTACTGTA CCTCAGCTGT TGTCCAGCTT CCAGTCTTGG GGTAATGGCA GTCTAATAAT	120
30	CTGAAAATTG CCAAGAGAAA GATGTGGAAG GATGAAATGG AGGCAACATG AATTTCTGTC	180
	ACCTTGTCAAT ATGTTCTCAT TTCCAKGCCT TGNGAGCAAG AGAGTTAGGT ATATCTCTG	240
35	TAACTCAGAC AATTTCTTC CTCTTTGCAG AATGGCCCT AGGAATCAAG GTAGCTTTTC	300
	TTTTGGAAAC TTICATGCTGT TTTTAGTGT GATAGAAAGG AGGTATCTGC CATTCTGTC	360
	ACCTATTTA TTTTGTGTA GCACCCATAA TAGATCAGCT GTCACAGCCA CAAATCTCTG	420
40	AGGAGACTGG AATCATTCCC AGATAAAATCA GAAAGTCAGA ATCACTTTAT GGTTATAGTC	480
	CTGGCTTCCTT GAGAGCTTGT CTGGAGGTTG TAGCAGGGGA GCACAGCTAG TCATATACCC	540
45	TWGACTARSQ ACCGGTCTWC CTCTATTGGG GATGGTTGTC CTCTTCTACT GAGCTTGCAG	600
	CTTTCGGAGG GACGCACATG GAGTGGTGAG GGAGGAAGGG GACACCCGCC TAGCCAGCCA	660
	GATCAGCTGA ATCAACCCCTG GCAATCAATG GGGTGACAGA TGTTGCAGCC AGATGCCCT	720
50	CACATCCAGT CCTACCTTCT TGGTAACAAA ACAATTGGTT TTGCTGGCT AGAAACTGTA	780
	GGGCTAGACA TGTATTATAG GACTGGCTTA GGGAGAGTTA CTTTATATTA GCACTCATGT	840
55	TTTCACTCAT TTATTCCTTG TAGCTCATTA AAAGAAAAAC CATAATTGAG CATCTACTAT	900
	ATGCCATGCA TTGTGCTGAG TATCCATGAT GCTCAGGTGA ACGGGACATG GTCCCTGAAA	960
	AAGTGTAAAG TCTGCTGGGA AAGTTAGTGC TCAAAAGTGT AACTAAATAC TTGAGGCAAG	1020
60	TGCTTTACTA GGGAAATAAC TAAATATCAA GAGAACAAAG ATAAGCAATT CCTTCACGAT	1080

GTTTTACATG GTAAATCCAT ACAATTTAA AAAAAAAA AAA

1123

5

(2) INFORMATION FOR SEQ ID NO: 57:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1239 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

GTATTGATAC	GAATTTGAC	TACATTTCTG	ATGGTGTGTT	TTGCTGGTT	TAACTTAAAA	60
GAAAAGATAT	TTATTTCTTT	TGCATGGCTT	CCAAAGGCCA	CAGTCAGGC	TGCAATAGGA	120
20 TCTGTGGCTT	TGGACACAGC	AAGGTSACAT	GGAGAGAAC	AATTAGAAGA	CTATGGAATG	180
GATGTGTTGA	CAGTGGCATT	TTTGTCCATC	CTCATCACAG	CCCCAATTGG	AAGTCTGCTT	240
25 ATTGGTTTAC	TGGGCCCCAG	GCTTCTGCAG	AAAGTTGAAC	ATCAAAATAA	AGATGAAGAA	300
GTTCAAGGAG	AGACTTCTGT	GCAAGTTTAG	AGGTGAAAAG	AGAGAGTGCT	GAACATAATG	360
30 TTTAGAAAAC	TGCTACTTTT	TTCAAGATGC	ATATTGAAAT	ATGTNAWTT	TAAGCTTAAA	420
ATGTAATAGA	ACCAAAAGTG	TAGCTGTTTC	TTAACACAGC	ATTTTACCC	CTNGCTCTT	480
35 CCATGTGGGT	GGTAATGATC	TATATCACCA	ACCTKAATCT	CTCTGCCTTT	TTTTCAAAC	540
ACCCCTTCAT	CATCCATCTT	AATTGCATA	AGGACATATC	TACTTTAATG	TACTACCACA	600
40 GTTACAGTT	AATGTGGAA	AGACCAGCTT	CACTATCCTC	TTCAAGCTAG	ATTCGCCTAA	660
CTTTTAACCT	TCACAGTTTC	CTGATTCAAA	TTTGCAGG	CTCTGATGCC	TTGAATTGGT	720
45 TTGCTCTC	TTTTTGGAT	CTGTTTTGTT	TGTTAAACAT	CATAATGCCAG	TCTCTCATTA	780
ATTTTTACCA	TCATTTACCC	TGATAATCTG	CCTCTTCTCC	ATTTCTCCCT	CCCTTACTAC	840
50 CTTTCTTGA	ATTTACTGTAA	CTGATTGGTC	CCACCAAAAT	TTTAAAGTAC	ATGAAGTATC	900
TTCATTGGTT	CATCCTCTTG	CCCCCTCCAG	ATGTCAAAAA	ACTTTATCCT	GCCCCCTAGC	960
TGACCCACCA	GGTTCCTTTA	TTTCAGTGCG	CCATGTGAGT	CTACCTTCCTC	CTAAGGAGTG	1020
55 CCCTAATCCA	GCCCTTTTTT	TGTTTCTTAT	GACCCATATC	TTTAGGCTCT	TCCCATTCT	1080
AGGTGGGAGA	TAGGTAAGTT	TCAAATCTAT	GCCAGTCTTA	TGAATATTAC	ATTAGGGTAA	1140
TGTGCTATAA	TGAAGAAATA	AAAAATACAG	TGCTTAAAAG	AAAATAAAAT	TCTATTCTG	1200
TCTAAAAAAA	AAAAAAA	CCNNGGGGG	GGCCCCGGT			1239

60

(2) INFORMATION FOR SEQ ID NO: 58:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 803 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

GGCAGAGGTC	AATCCAGGAC	TACAAACACC	TGTGCCAAGA	CCTGAGCTTC	TGCCAGGACC	60
TGTCATCCTC	CCTCCATTCTG	GACAGCTCCT	ACCCACCGGA	TGCGGGCCTG	TYTGACGACG	120
15 AGGAGCCTCC	CGATGCCAGC	CTGCCCTCCTG	ACCCGCCACC	CCTTACTGTG	CCCCAGACGC	180
ACAATGCCCCG	TGACCACTGG	CTGCAGGATG	CCTTCCACAT	CAGCCTCTGA	AGGGCTGGGG	240
20 GCCAGGGGGC	ATGCACCCAT	GCAAAAGGCT	CAGAAACTCC	CCCTCCGGCA	AGCCCTCAGA	300
CTTCGGAGCC	TGCGCCTTCC	CCCCTACCGC	CTCACCTCAC	AGGAGGGCCA	GGCATGTATT	360
25 CCTCAGAGGC	GAAACTGCCA	AACTCTTTCT	CCTGTCTTGG	GTGGCTGGC	ACTGGGGCGG	420
GCATCTAGGG	TACAGCCTCT	GTCATGGCA	CTGGGCCTCC	AGTTCTTCCA	CATGTGTGCA	480
CCCCAGCTT	GGCCAACCCCT	CAGCCTTGCG	GTGGGGCCCG	AAGCATCTTC	CTTCCCGCTT	540
30 GCGCTCTCTG	GGATGGGAT	GAGTGCTGG	CTCCCACCTC	CTCCTCACCT	TTTGTGTGTA	600
TCGGCAGCTG	CTGGCTCAGG	GGCATCCCAM	CTCCGGCTC	TGGGTTCTC	TGCCCTGGAA	660
35 GGGCTCCAGG	ACCCGTCCCA	ATAACCACCC	ACGGCCAGKA	RGCCAAGGCC	CCGTGCTGGA	720
TATTTAAATT	TAGGGGCCGG	TCTCCAGGGC	GCGTAGATAA	ATAAATACAC	TCAGCGTCAA	780
40 AAAAAAAAAA	ARAAAAAAA	ATT				803

(2) INFORMATION FOR SEQ ID NO: 59:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 995 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

GATTCNGCA	CGAGGNAACA	GCITTATTCT	TGGTTATTCC	TAATGTCCAC	CTAGTCCTCT	60
55 TTWACTTYC	TTGGTAGGGT	TAGGGTGGCA	TGGGGAAATG	GGACGGTATC	ATTTGTCTT	120
TTTAACCTTT	TTTTTTCCA	CCTACAGCAG	CTGTTTTAC	CCTGTGGTCA	GTCAGGTACT	180
60 ATATTTAGTT	TGCAGTTGCA	CTGCTGATCG	ACCCCTGATG	GCCCCAGTTG	GAAGTTGTTT	240

	GGGGGAAGG AAYTAGGAGA GGCCAGGSOC TCCATTTAAA CCATGTCGT AATGTCCT	300
	TGGAAAGAAA AAAAGATACT GTTCCAGTCA TGGTTTCTTG GTAGTTCAGC TTTAAATGG	360
5	GCCTCATTAA AAAATTCAA TAATTCAGGC TAATTTTTC CCTTTATATG GTAACTCCAC	420
	CAAGTTGTC TAAATGTATG ATTTTTATCA TGATTAAGTT TTTAYTTCCA CATCATGTGA	480
10	CAACTGGCCT GGGATGGGAT ATAAGCTCAG AACACAAAGT CATTACCTC TTAAAAAAAT	540
	AATTCTATCT GTGGGGGTT ATGTTATTTT TGTTCAAAGA GGACACAATA TGATGCAGAA	600
	TACACCATG AAGGATTTT TGGTTGGCA AGTTCTTATT TTTTTAAATG CCTGTAAAAC	660
15	CTAGCAGTGT TTCTGAAATT GCATACCTTA CCTGATGTC AGAGATCCGA TTTACTTCCT	720
	GATTTCCAG CAAGTGAATT TGAAAACATT TAATCTAATC ATTCCCCCA CCGTCTGTT	780
	AAATCAAAGG AAGTGGCATC CAGCACTAAT TTTCATGCAT TTATGAAAGG ATGCCTGAGG	840
20	ACCCCTAAGT ATAATTCAA AATTTGTTA ATGTGTGTC CTTGATGAAG TTCTTTAGGA	900
	GTCGTAGAAC GAACTGATTG CCCACTGATC ATCAAATGCA AGTTATGAAC ATTTAATAAA	960
25	AATTTAAAC CAAAAAAA AAAAAAAA CTCGA	995

30 (2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 966 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

40	GACACTACGG TCCGAATTCC CGGGTCGACC CACGCGTCGG GGAGAGGACA TGCAGTGGC	60
	ACAGAAAGTT CAATGGAACA GATGCCACTG TGGGCACCAA GACTGTAATG ACTCTGTG	120
45	GTAGGTAGTT TAAAGGACT GCATGCCTTG GAAATGATTTC TTCACTTGGA GAACATACCT	180
	GCCTCTAGAT ATGTTTGTC CTCCTAACAT CCTGAATATA ACAATAGAGA AAGATAAGTC	240
	AACCAACAGA TTTAGGGATG TGTTTCTTCA GCACATTTG GTCATTTGA TGCCAAAGTT	300
50	GACACTGT TTAATTGGC AGCACCTTG CTCCCTTACCC AGGTATGTAT CACTTGTGA	360
	CTCCAGGTGC CATTCTTGGT GATGACAGAA TGTTTATCAC TATGTTGTT AGCAAGAGGA	420
55	AGCTTCAAT ATAGGAACCT AACATCTTCC CATGAGTATA AATGAATTAA AGACATTTGA	480
	ATCAAACCTT CAGTAGAGGG AGGTTTTAGA ATTCAAAAA CTGGTTAAG GAAATTCTTT	540
	TTACTTTCC CAAGGTTAAT CTTTTAAAT ATCTCTAGAC ATCAAATACT TTCTGTATGT	600
60	ATTAGCTGTG TCTGTCTATG ATGCAAGTAA CTCTCCTCCT ATTTGGGGA TAGTTCAAG	660

	AGGTAGGAGC ATTATCTCCC ATTTTCTGG TGACTTCTTG GAGTATAGAA TTCACCATT	720
5	TATCCGTAAG TCTTCAAAGG ATTATGGTGG ACTAGAACCT ACATAGTGCA AAATAGTCCT	780
	CTATTTTAA TAGGAACCTA GAAAAAACTT AGAATTATAT ATAGAGTTGT TTCTTTAGA	840
	AACCAGAGCT ATTATTTTGT ATTTAAAGCA CTGTTTATTA TTTGTACTGA TTCTTATCCC	900
10	TCTGTGTGAA TAAATGTAAG ACGGTGAAAA AAAAAAAA AAAAAAAA AAAAAAAA	960
	ACTCGA	966

15

(2) INFORMATION FOR SEQ ID NO: 61:

	(i) SEQUENCE CHARACTERISTICS:
20	(A) LENGTH: 262 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

	TTGCAGGTAT ACATCCAGAT GCACAGAATG TCCATTTGTC CCTTATTGGT GATGCTAATT	60
30	TTGATCACTT GGGTAAGATG TCCAGTTTCTT CCAGTGTATC GTTATTGTTT TTCTTTTGC	120
	AATTAGTGGG TAATTTGTGA GGAGAAACTT TGAGACCTTG TTTGACAATT CTGTTCCCTCC	180
	ATCAAATCTA CCCCTCCCTA GGTTTAGCAT CCTTGACAA TCCTTGTCT GAATAAATT	240
35	TTAACTAAGA TGTTTNCCCCA AN	262

40 (2) INFORMATION FOR SEQ ID NO: 62:

	(i) SEQUENCE CHARACTERISTICS:
45	(A) LENGTH: 753 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

50	GGCACAGGT CTTTGCCAG TCATGACAGA ACCATGCAAG ATATGTTA CAAATTGGTA	60
	CCAGGCCTCC AAGAAGGTGA GTGTCTGACT GTCTTGCTGA TCCCTGAGGT CCCAGCCTGG	120
55	CCTCTGCAGC CCCTGCTCTC CTGGAAGTTT GGTTCTCGGA TGGGAGGCC CTTTCTTTTT	180
	GGCCGAATCA CGGTCTCTC ATCCCTGCTC TCAGCCCAAC TTCATCTCCT TGGCTGGTCT	240
	CTTCTTTCGT CTAAGATGCG TAKACATCTT TTTACCCCTT ATGTGTATTTC ATTCAGCAAG	300
60	TATGGATCGC ATGTTTAGCA CATGGAMCC CCAGGGNTCA ACGCAGCTCC TGCCCTCCCC	360

	AGGACCCCTGC CTTSTTCCTG GCCCCCACCT CCTGTCCAG GCCTGCCCTCC CCTCATCCCA	420
5	CAGGCCAGC TTCCCCACAA CAGAGGAGCA GCACGTTGGC ATAGCGGGTA CCTGGTGTGTT	480
	CTAGAAAAAC TTCACCATAA AGTCAAATTT CATTAGAAT TAAAAGAAAT ACCAAGTAGT	540
	ACAAATACCC TGAAAGTGGG AATCGGTTGC TTGGGGATCG CTCAGCTGAA AGCTCCCCCA	600
10	GCTCCCGACA CTCTCACGGT GGTGGCCCT CCGCTGGCGA ACCGGCAANG AAGCCCAAGG	660
	AAGGGGCCA GGTCAGCGC CCAGGTTGGG CTTGTCCTG GTTATTCTG CTCCATCCAN	720
15	AACCTTTCCA AAAGGCAGAA TAGAAAAACN TGA	753

20 (2) INFORMATION FOR SEQ ID NO: 63:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 739 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

	ACAATACATG CATCATATCT TTTGACTTTG AAGGATATCT CATGTCAAAG GAATCAAGTT	60
30	ATGATTTATA GAGGATTCAAG CTGGAATACC TTGTGGGTGC TGGCTGAGGG TGGCAAAACG	120
	CCTACCGAGA CATGAAGGTT TTAGCCACTA GTTTTGTCCCT TGGGAGCCTG GGGTTGGCCT	180
35	TCTACCTGCC TTTGGTGGTG ACTACACCTA AAACACTGGC CATCCCTGAN GAAGCTGCAA	240
	GAAGCTGTGG GAAAAGTTAT CATCAATGCC ACAACCTGTA CTGTCACCTG TGGCCTTGGC	300
40	TATAAGGAGG AGACCGTCTG TGAGGTGGGC CCTGATGGAG TGAGAAGGAA ATGTCAGACT	360
	CGGGCGTTAG AATGTCGTGAC CAACTGGATC TGTGGGATGC TCCATTTCAC CATTCTCATT	420
	GGCAAGGAAT TTGAGCTTAG CTGTCGTGAGT TCAGACATCT TGGAGTTGG ACAGGAAGCT	480
45	TTCCGGTTCA CCTGKAKACT TGCTCGAGGT GTCACTCTCCA CTGACGATGA GGTCTTCAAA	540
	CCCTTTCAAG CCAACTCCCCA CTTTGTGAAG TTAAATATG CTCAGGAGTA TGACTCTGGG	600
50	ACATATCGCT GTGATGTGCA GCTGGTAAAA AACTTGAGAC TCGTCAAGAG GCTCTATTTT	660
	GGGTGAGGG TCCCTCCTCC TAACTGGTG AATCTGAATT TCCATCAGTC ACTTACTGAG	720
	GATCAGGACT AATAGAGAA	739

55

(2) INFORMATION FOR SEQ ID NO: 64:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 476 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

GAATTCCGCA CGAGAGGACA TGGATTATGG GTACTACTCA GCAGGCCAGT TTTTACTCCA	60
10 CCTCTTCTA GCTGACTTGA CACAAGCAAC AACCCAACAG AAAACCAATA CTTCTGAGAA	120
TGGCTGCAAG TTTGTTTG TGCTGCTTTG AGCTAAGAAA TCAAGGCTGA GCTCTCTT	180
15 CTCCCTAATTG TCAGGAAGGA GGAAGGCAGA TGTTGAGAACG CTGATTGGGT CTGAGTGTAC	240
TGGGCAGCAT CACTGTTAAA AGGTTCAGCAC ACAGATGCAA GCTCACTTGT CTGCTTNCCTT	300
TCATGTGACT GAAGTGGTTA AGAARGTTGT NCAACTCCCC CCTGCACCCC CCTCACCCACC	360
20 GCAGTAAGGG AGAGACAGGG CCAAACCTGC AGCTTCGGTA GAAGAGGCCA AGGCAGGTGT	420
CCAAGGCCAG ATCAGCAGTC AGCCAGGGCA AATGGGCTCA CTCTGGTTAC ATGACC	476

25

(2) INFORMATION FOR SEQ ID NO: 65:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 754 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

AATTCCGCAC GAGACCAATT GTACTTTAT TATATCAGGC TGATTCACTG TTTCTAATGC	60
40 AATGAACCTTG ACACAGATTT TAAATTTTTT CTCAAATCTGT CCCATTGTGT AGACAAATTA	120
ATTCAAAGTT CTTTTCTTC CTTCCTCTTT TCATCTAACG CTGTCCTTAT GAGTAGAAAA	180
AGAGAAGAGG CTACCTTGAA ATGCCTCGGG CCCAAACTCA GAAGGCTCTG CACTCAACTG	240
45 AGCCTCCCTT CCTACTAAGA ATGGAATAGT GTTGCTTATA GGGGTGTGG TCCAAGTATC	300
AGCTGTGGAT GATTAATTC CAGGGCTGCT ATCACCTAAG GTAACTTCAG TAATCTTATG	360
TGTTTGGAAA GGAGGATGAG GATTATTTT CAAATACATA ATTTTGTGTTT ATTTTGAAC	420
50 AATCTCACAC CTACAGAAAA GTGCAATTAA TAATACAAAG AGCTTCCCCC TCGCCTGAAC	480
TGTTTGTAG TAAGTTGCC AAACTGATAT ACCCACGATC CCCAAATGCT TCAGTGTAT	540
55 TTCCCTCCAG CCAAGGACAT TCTCCCTGCA TAACCCACAA TACAACCCAT AAAAGTCAGG	600
AAAATTTAAC ACCCAGTTCC ATTTTGAAC CCATCCTGAA ATTCCAGGTG TTCATTCCAT	660
60 GTTTTGGCC AGTTGGTNCC TTGGTATGT TCCCTCCNT AGCCAAAAA AAAAAAAA	720

AAACNCCAAG GGGGGGGGCC CGGGTCCCCA ATCC

754

5

(2) INFORMATION FOR SEQ ID NO: 66:

	(i) SEQUENCE CHARACTERISTICS:	
10	(A) LENGTH: 1890 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:	
	GGCAGAGRAA AAACAAAATG GGTAAATGCAT TCGAGGTGAC AGGGTTAATG TTGGCATTAC	60
	TTTGTATGT TGTGATGGG CAGAAACCCA AGGKGGGGTT TTGTTGAGCA TAAACACAAG	120
20	AAGCAATTAT TTGTCGCACT AGACTTAACC CAAAGGACAG ACCCCTACAT GTATATAGTA	180
	GAGAAATCCT GTCTTTAGC ACTATCTCAC AGGGGAAGCT GAGGAATCAC ATTATCTTA	240
	ATATAAATAA ATGAAATGCN ACCACTGTAT AATTTATATC CTTAAGCAAC TGGATTCAMC	300
25	GTACCACTAA TGGCCTGGTC ATGTTTTAAA CATTACCCCA AAACAGCCTA ACTGTTCTGT	360
	GACTCAGTGT CTCTGTGGAA TCCTATTTAG TAGCACCATG GTCTCTAAAT GTTTGATTA	420
30	CACATCAGTA TTAGGAAAAC ATGTTTGAAG CATTGTCTAA GTCTGTTTGT GCTGATGTAA	480
	CAGAATACCA TAGACTGGGK AGTTTATAAA GAGAGAAATT ATTGGCTTAC AGTTGTTGGAG	540
	GCTGGAAAGT CTAGTATCAG CGTACTGGGA TTTGGCAAGG GCCTCTTGG TGCATGATAG	600
35	TATGGTGGAA GGTATCACAC GGCAGGCAGA AAGGCAGAGA GAGAACAAAA GGGGGCGAAC	660
	CCACTCCCTT GATGAGAACCC TAAATACCTC TTAAAAGTCC TAACTCTCAA TGCTGTTTAC	720
40	AATGGCAACC AAATTTAAAC AAGAGTTTG TAGGAAACAA AACTCAATC AAAACCATAG	780
	CAAGTATGTA CCATGACTGT ATGIGTATTT ATAAAATACA TTCAATATATT TCTACAGCAA	840
	TATATATGAG GTACATTTAA GCATGTAAAA ATAGGAATT TTAAAAATAG GACAGTTGTA	900
45	ATAATTTCTT TGACATTCC ACTTTGGAGA CTGTTTTAT ATGGRGCTTG TTTTATCACC	960
	AAAAGGCATT TTAATTTGTC ACACTTAGA WTTCTTACAA TGTGTAATTG ACTGCTAGTT	1020
50	GCTGAACAAA GGACAGATAA AGTGTTCCT GCACCTGAGC AGCCTAAAGG TGAGTGTAAAT	1080
	ACAGATGCAC AAGTGACTGG TTGATAATGG AATGAGACCC CTTATAAGAA AGACATACAG	1140
	AGCACGGCAG AGGAGCAAGA ACMACACAGA GGCAANTGACA TTTGAGCTAG GCCTCTTATA	1200
55	TCTGTAGATG AACATTTGAT GGTAGGTAGT AGGGAAGATG GAACTAAGAA TATTTGAGCT	1260
	ACTTAATATA TGCCAGGCAG CATGCTGAGT GCTTGTGTC ATTAAATTCT CAAGACAGCC	1320
60	ATAAGCGGCA ATACAGGTAT TGGGCCTATT ATTCTAAATC CCATTTATAA AGAGAGTTAG	1380

	GATTAGATTG AGTTCCATCT TTCTACAAAA CCTGGCACTG TCATTCCAGG CAAAGGGAGT	1440
5	ACAATCCATT TTTCTCTTAA GAGGTTGATT TTGCCAATGA GACAGAATGA ATCTCTACAG	1500
	CTTGTAAAGT TTCWACCGT CTTTGGGTGA CTGAAAAATT CAAATGTAAA GATGTGGCAA	1560
	AAATGGTTCT CTAAGGATTT TAAGTACAGC CAAATGATAT GTCACAAGTT TTTTCCTAAA	1620
10	TATCCAACCA TTTAGCTTTT CATAAGCTTT TAATTCCACT AGCCTCACIT TCTGAGATIG	1680
	TTGATGTTTT CTTGTCTAA CCTGAAATTT TCTTGTGTTG ATGTTAACAG GAGTATAATG	1740
15	AAGGAGTAAC CATTTTTATT TTATGATAGT CTATCAATAG ACTTTTTTTA ACCTTCCTTA	1800
	AGCTAGGTGT GTTGTCCCT TATTAAGTC AGTTTGACCC AGCCTGTACA ACATTGCAAG	1860
	ACCTTAACCT TAATAAAAAA AAAAAAAA	1890
20		

(2) INFORMATION FOR SEQ ID NO: 67:

25	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1614 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:	
	AAATAAGACN TCTTTGAGCA GCGATTGCTG GATCATTGAT CTGTTTGAGG AATGTCTGAC	60
35	CTGGCCCTRA RAGCTGGAGA AGGTGCAGAT TCAAAGTRAG CGGCTCCTRA GGAGAGCCCC	120
	AAGSTGCTCG CCTTCCTCCGT GCCTTCGGCA GCTACCGCTCT GCACGGTGAG AGGGCACGGG	180
40	CACACGGTTC GGGCTGGCGT GCAGTCTCCC AGCCAGCCAC GCTCTGCTCA GGCTGGAAG	240
	TGAAAGCCGC CTCCCTCCCG TTATGCCCCC CATAACAGGAG CCTCGGTTTT TCAGCAAAAC	300
	GCGGCCAGTC CCCTTCTCCA CTGCTGCCTC CCAGCAGAGG GCCCCAGGAT CTCCAAGGTC	360
45	CCAGCTATGG CTTTGGACAA CGTGGCTTCG GCCCCCTGGGG TTGCAAGAGCT TGCAATTGGT	420
	TTACCTCGGT CTCATTCATT CATGGAGCCA AGGGTGGGGT TTCACCTGCG AACATCAGAC	480
	TGACTTGCTG GCGTCAAGAG CAGTTGACTC ACTGATGAAG GCCCCTGGTGA GGAGAAAGCA	540
50	CTCTGTTCTT CGCCTACTCT GTAAATGTTT TGTCTATAATG AGCCATGAAA AAAGTAATGA	600
	ACTTGTGCTG TTAATCGTCA CTGTAATGAG AAGTCTTAAG TACAACATAG CTGTGGTGGC	660
55	TGCGTGGTTT AATGGCTGCA TTAGATAGGA TCCTCACATC CCATTCAAGAA CCAAAACTGA	720
	TACAGTGAAA CAATTAAGGT GAGCAAATAG TTTTAACCTTT TCTTTTTTTT TTTAAGTTTC	780
60	ATTCTTCCTA GAATATTTT CTAACAAATT TTATTTCAAGC TTTAAAGATG GGTCAATATAG	840

	CCAAACGGGC CATATAATCC AACATTGTTG AGATGTCTTA GGACATCTAA GGCAAAACTG	900
	GCACATTGT TCTGCAGACT ATTGCAGGA TGTTTTTCTC TAGCATTTCT ATATTATCTG	960
5	TCCATTCTGA GGAACCCAGTG AATGTCCTAT AAATGCACCT CCTGTCAAAA CCATGCCTGA	1020
	GAGGTCCCGG CTGGGAGTGA CAGGGTGCCTT NCTTAGATTC TATTGGCTCT TCTCTCATTC	1080
10	TCCGAACCTTA CTCCCTTTTA TGGGTAAGTC AACTAGGTYY ACAGTCCCTT ATTTTTAATG	1140
	CCTAAGTTT GACAGCAGGN AAGAAAACAA TTTTTAAAAA ATTCTCATTA CATAGACGCA	1200
	CAAGAATATG TCACATAAAAG AAAATGTGTT TAGAATACTG GTTTTCTATT TACGCATGAT	1260
15	ATTTTCTAA GTAAAATTGC CAAGTGGACT TGGAAGTCCA GAAAGGAAAA TAATTAAAT	1320
	TAATGCTGGT GATCTTAACA ATATTTTGTA AAATGATGCT TCCCCCTCT CCATGGTGT	1380
20	GTCAATTTCG TACAATTAGG TATCTGACTT TACAAGTTG TTATCCCTTC TAATTTTAC	1440
	TGAACtgAAA GCACAAAGAA GACTACACAG AAAATCTGGA AACAGTTGCA GGTGTTGGGA	1500
	GGAAGATGAA ATCGAGCTGT CTTTTAACCT TCGTATGTGT TTTATCAGAA TTTGCTGGAC	1560
25	TATGCTAGCA AGGACTTTGT TTACNATCAA ATTGTACTAG TGTCTGCAGG GTTT	1614

30 (2) INFORMATION FOR SEQ ID NO: 68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 596 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

40	CTTTTCACCC TTAGAGACAG GGTTTCACCTT TTTTGCCTTC TTAATGGAGA TATTCAGTTT	60
	TCTTTTTTTC ATTTAAACAA AGAAAAAAA TGTATCTACT CTACCTTCCC TCTGCTCTCC	120
45	TCCCTCCCTA TCCTACTTGC CCATATGAGC ACGGCTCCCC ATGCCACAT ACTCCTGCAA	180
	AGCTTTTATG CTGCTTCGCT TTTCTCTAAA CAGATCTGAT ATTGCTGCTC CTGTGGTTTT	240
	CTCAAAATTA ACTTTGCCGT GTTTTTAAA AAGGAATCAA AATGCATTGT TGCATTAAGC	300
50	TTTTCAATA AAGGAAAATT ACGGAAGGAA AATAGGCAAC ACCAGCAAAT TATATGTGGA	360
	CAGGTTCTAA ACTCTATATA TACATATATA TATATATATC TATATATCTA TATACGTAAT	420
55	CATCTAGTTGTC TGTCTATCTTA CTGAAAGGAA TAACACTTCT AAAGATCACC ATTTCTGAGA	480
	AGTTCTTGGAA ATCTTTATG TCTAAGTGAT TGTATTAGAT CAGCAATAAT GACTATGTAA	540
	TCTCAAAAAA CAAATAAAAT ATTCTTAACA TGGAAAAAAA AAAAAAAA ACTCGA	596

(2) INFORMATION FOR SEQ ID NO: 69:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1524 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

ATCCGGAATT	CCCGGGTGTG	TTCGACCCGT	CCGGGACTTT	GCACAGCACC	TTCCAGCCCA	60	
15	ACATTTCCCA	GGGAAAACTT	CAGATGTGGG	TGGATGTTTT	CCCCAAGAGT	TTGGGGCCAC	120
	CAGGCCCTCC	TTTCAACATC	ACACCCCGGA	AAGCCAAGAA	ATACTACCTG	CGTGTGATCA	180
20	TCTGGAACAC	CAAGGACGTT	ATCTTGGACG	AGAAAAGCAT	CACAGGAGAG	GAAATGAGTG	240
	ACATCTACGT	CAAAGGCTGG	ATTCTGGCA	ATGAAGAAAA	CAAACAGAAA	ACAGATGTCC	300
	ATTACAGATC	TTTGGATGGT	GAAGGGAATT	TTAACTGGCG	ATTTGTGTTTC	CCGTTTGACT	360
25	ACCTTCCACC	CGAACAACTC	TGTATCGTTG	CGAAAAAAAGA	GCATTTCTGG	AGTATIGACC	420
	AAACGGAATT	TCGAATCCCA	CCCAGGCTGA	TCATTCAGAT	ATGGGACAAT	GACAAGTTTT	480
30	CTCTGGATGA	CTACTTGGGT	TTCTCTAGAAC	TTGACTTGCG	TCACACGATC	ATTCTGCAA	540
	AATCACCAGA	GAAATGCAGG	TTGGACATGA	TTCCGGACCT	CAAAGCCATG	AACCCCTTA	600
	AAGCCAAGAC	AGCCTCCCTC	TTTGAGCAGA	AGTCCATGAA	AGGATGGTGG	CCATGCTACG	660
35	CAGAGAAAGA	TGGCGCCCGC	GTAATGGCTG	GGAAAGTGGA	GATGACATTG	GAAATCCTCA	720
	ACGAGAAGGA	GGCCGACGAG	AGGCCAGCCG	GGAAAGGGCG	GGACGAACCC	AACATGAACC	780
40	CCAAGCTGGA	CTTACCAAAT	CGACCAGAAA	CCTCCCTCCT	CTGGTTCAAC	AACCCATGCA	840
	AGACCATGAA	GTTCATCGTG	TGGCGCCGCT	TTAAGTGGGT	CATCATCGGC	TTGCTGTTCC	900
	TGCTTATCCT	GCTGCTCTTC	GTGGCCGTGC	TCCTCTACTC	TTTGCCGAAC	TATTTGTCAA	960
45	TGAAGATGT	AAAGCCAAT	GTGTAACAAA	GGCAAAGGCT	TCATTTCAAG	AGTCATCCAG	1020
	CAATGAGAGA	ATCCTGCCTC	TGTAGACCAA	CATCCAGTGT	GATTTTGTGT	CTGAGACCAC	1080
50	ACCCCACTAG	CAGGTTACGC	CATGTCACCG	AGCCCCATTG	ATTCCCAGAG	GGTCCTAGTC	1140
	CTGGAAAGTC	AGGCCAACAA	GCAACGTTTG	CATCATGTTA	TCTCTTAAGT	ATTAAGGTT	1200
	TTATTTCTA	AAGTTTAAAT	CATGTTTTTC	AAAATATTTC	TCAAGGTGGC	TGGTTCCATT	1260
55	AAAAATCAT	CTTTTTATAT	GTGTCCTCGG	TTCTAGACTT	CAGCTTTGG	AAATTGCTAA	1320
	ATAGAATTCA	AAAATCTCTG	CATCCTGAGG	TGATATACTT	CATAATTGTA	ATCAACTGAA	1380
60	AGAGCTGTGC	ATTATAAAAT	CAGTTAGAAT	AGTTAGAACAA	ATTCTTATTT	ATGCCACAA	1440

CCATTGCTAT ATTTGTATG GATGTCATAA AAGTCTATT AACCTCTGTA ATGAAACTAA	1500
ATAAAAATGT TTCACCTTTA AAAN	1524

5

(2) INFORMATION FOR SEQ ID NO: 70:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 819 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

GGCACGAGGG AGAGGGACGG GGAGGGGGCG AGGGGCGGAG GCCGAGGGGG CAGGGGNTGG	60
20 GCGCCGGCCA GTGTTTACAG ATGAGCTTTA ACTGCCCT CAGCGTGGA GACGGAGACC	120
CCCGAGCCCG GCGGCGCCTC AGCCCTTCAA CGACAGTATT GAGTGGTCAG GTTACAATAA	180
25 ACCGGAGAGA AAAGGTCCGC TTGCACTTTT TTAGTTTTC TTATTTTTAG ACACCCCTCC	240
CCTCCAGGGT GATCTTTAAA AAAGCAAAAC AAAAACACG ACTTTTCCAG CGCTCAGCGT	300
TTTTTCCCTTT CGTCCGAAGC CGTTTTCTGA TTGACTTTT CTGGCGGGCC GGTCAGGCG	360
30 CCACAGACGT TCCAGAGGG GAGGGTGACA TTTTACTCC CTTTTGGGG CTAACCATT	420
ATGCTTTTGT ACATCAACCG TGGCGGGCCG GAGGGGCAG GGGGGGGGG GCGAGGGCG	480
35 TTCCAATCAA ATTTCTAATT TCTGTTAATT ATTAATCCCC KTTTACTGC GGTTTCTGTT	540
GTCATTTTTA AAATTTTTT AATTTTTTT TTTTTTTTAC TTTTACTTTT TACCTCTTGT	600
GTATATGTAG GGAATTATA GGGAAATATG TACTTATGG AATAAATTAA AAGAACTAAA	660
40 ATATATTATA TTTAAATAA AGTAATGGAC CTTTAATCTT ACACAGCTAA ATTACTGATT	720
ATATATTATSC TGAGCTGATT TAAGGGTTAA AAAAATTGTA TCAAGAGTTT TATTTTTGA	780
45 CTTCAAAGCC TTCTTAATAA AGCCTCTTTT CTACATGTG	819

50 (2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1442 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

60 AATTGCTTGG CATGAGTTTA CTTTAATGGC TGTTTCTGAG TTTGATCCCT CTCCGGAACC	60
----------------------------------------------------------------------	----

	AACCSCTCTG ATGTGTCTTG TTCCAGCAGG AAGAGACAGA CCTGGAGGTT CTGTACTTGT	120
	GATTCTGGT TGTGGATCCT GAGAACAGA AGTACTGGG A CCTAAAGTT CTGACATTG	180
5	CAAACAGAT TAATGACCTA CCACATTCCA GATCATTGG TGAYYWTGTG TTGTGCGTGT	240
	GGGTGTGTG CTTGTGTG CAAATTCAAG GTGGTCCCAG CCTTTCTAGT CTTCTCTAAC	300
10	CTTTCTCTC ARAARTCGCA CCTGTTCTGT CTTCTAGGA TATAATTTTT TTTCTAATTAG	360
	CCTGGTAAC ACCCCAACCA ATAAAGTTG CAATATCCAA GCCTCCATAAT TTCTCTACTT	420
	ATTAAGCTTC AGCATGAGCA AGCCTAAAAA CTCGCCATTA TCTGGAAAAG	480
15	TTCTATTCA CAGGCTTTAA TCTCTCTAG AGTAGTTAGC ACTCTTTGTG GGCTTTGTGT	540
	TCCTGTACTA GCTTGAATT CACAGTCTGA CGTTAATAAT TAGCTCCCTA ACACGTCCAT	600
20	CCTCTCTTGA TGTCCTGCTC TCTATTTCCTT CTTCTTTCTT CCAAGTTGGG ATAAATTCA	660
	CTTCTTATTTC TCC TGCTCCA GAMCTGGTT GTGGAGAAAG ATAGAAAAAG TTCCATACAG	720
	GGGACTCTGT GATCCTGCTA ACATCATTAT TTACCTAACG TCTTTAGACT CCAGTGAAAG	780
25	CTTCTGATT AATGTCTGTG CCCTACTTTA TGCCACATGT CCCATACCAT TTTCTTTGTT	840
	TTATGCAATT TATTTCCACT ATCTGATCCC ATTCCACCCA CATGACTTTG AGTGGAAAAC	900
30	TTCATCTCTT CATTGCTGAG TAAACAAACT TCAGGATGAA CAAGCCCTGT CCACATT	960
	CCCTTTACT KTAARKYCT GGAATTWWA TGATCTACGT TTTTTCTC TGTTTTATT	1020
	CTTCACTCCA TATCAACTTA CTTGGGGATC TACACCTTCA TTCATCTTTC TCATTCTGTC	1080
35	GGCACCTGGC TATGGAGTTT ACATTTCTCA TCATATTAC TCCTCATAAT AATCTGTGA	1140
	GGTATATACC ACTCTGAGTC TTGTATAAGA GAAAAAGAAA CTGAGATAGG GATAACTCAA	1200
40	AGGGATAATT CATTGCTGG AGCTACCAAC TAGCTACTAA CCATGCTAGA ATGGACAGAG	1260
	ATGACATICA TGCCAAAGAC CAATGTTGACT TGCTATCTC ACATTTGCTC TAAGTTAGA	1320
	AAAAAAAAT CCCTCAATT TATCCTCCAA CAGCTTCTT AGAACCTTAC CATGGATGCC	1380
45	TTGTWTAAACA CATTTCACCT TTCTGGTAAA AAAAAAAA AAAAAAAA AAAAAACTC	1440
	GA	1442

50

(2) INFORMATION FOR SEQ ID NO: 72:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1223 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

	AACCTGAGGA GGCTGTCATG ATAGGAGATG ATTGCAGGGA TGATGTGCTT GGGGCTCAAG	60
5	ATGTCGGCAT GCTGGGCATC TTAGTAAAGA CTGGGAAATA TCGAGCATCA GATGAAGAAA	120
	AAATTAATCC ACCTCCCTAC TTAACCTGTG AGAGTTTCCC TCATGCTGTG GACCACATTC	180
	TGCAGCACCT ATTGTGAAGC AATGTGTGCA TCTGAAGCAA CTTGAAATGC AGCTCTTAT	240
10	TGTCTGGAAT GAATCCCTTA CCAACTCAGT GCCAGCATCG GTAGACACCA GTCACTGCTG	300
	ATCGCTTTTT AACCCCTCTT TGTGTGCAT TAATTAGAAA GAAAGGTATT GAATTGCGGC	360
15	TAGCCAGTAA GCCTTGCTAA TCTCTTTAT TTGTAACTG AAGATGAGAC CCAAAGAAAG	420
	GGAAAGCTGA GATTTTGTGC CATTCCCTTT AAAAATTCA TCAGGTTAGG TGGGCTGTG	480
	GGGGAAAAGC TACTACAGGG AAGAGTGTTC TCTGCTGTCT CTTCACTGGA AAACAGGGAG	540
20	GGGGATTTTC AGACTGTGAA GAAAGTTGAA TGGTGGTTTT TAAATTATAA AGTAATGTAT	600
	TAAAAGGTGC ATTAGGCTGT AGTTCTAATA TTGAGTTCAA CTGTGAAATC CATCAGATGT	660
25	GCCAAATGGA GAAGACAGAA ACCAACAAAG TGAATTGTTC TTAGGCCAA GTGGTACAGT	720
	GAATTTGCTT TAACAGATGT TGAAAACCTAA ATTTTCTACT GTATTCCAG CACGGGTGAC	780
	TTCTTTTCT CTCATTAGC CAGAGATGAC TAATTAAAT TTAGAACCAAG ATTTTAATT	840
30	AAATTAATAT TTCCATTAAAT AACCTATTCA TTGCAGATAC CTATTATACT GTGTAACAGT	900
	TGTTTGGAA ATTATTATGTA AAATTAAAAC TATCAGTATT TTACAGATGT TTAAATTAGA	960
35	CATGTTATTAA ACAGGAACAG TGCAGAAACT AGAATCAAGC CTTATAATAT CTTATAGACC	1020
	ATGCATTTTG AAGTTAGTGT CCTATTAACT GTACATTGCA AGATTCTTA	1080
	TTTGCCCTCT GACACTAWGG GAAAATTTT AGAAGCCAAT GGGACAGATT CCAGCCTTTA	1140
40	ACCACTGGGT ACTACAGCCG TAAAAGGAAA TCCCGCCTGG TAGCCAGGGA TATNCCTCCC	1200
	CAGGTAAAN CCCCCCAAAT NAA	1223

45

(2) INFORMATION FOR SEQ ID NO: 73:

50 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1814 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

	CAAGCTTGT ACTTAGATCT TTTACTTAAAGA TCTGCTTTT GTCTTATTCT TTTTAGTGGA	60
60	TGTTTCCAAG GATTTGTCTTC AGTCATGGCC TTGGGATTAA AGTGCCTCCG CATGGTCCAC	120

	CCTACCTTTC GCAATTATCT TGCAGCCTCT ATCAGACCCG TTTCAGAAGT TACACTGAAG	180
	ACAGTCCATG AAAGACAACA TGGCCATAGG CAATACATGG CCTATTCAAGC TGTACCAAGTC	240
5	CGCCATTTG CTACCAAGAA AGCCAAAGCC AAAGGGAAAG GACAGTCCC AACCAGAGTG	300
	AATATTAATG CTGCCCTGGT TGAGGGATATA ATCAACTTGG AAGAGGTGAA TGAAGAAATG	360
10	AAGTCTGTGA TAGAAGCTCT CAAGGATAAT TTCAATAAGA CTCTCAATAT AAGGACCTCA	420
	CCAGGATCCC TTGACAAGAT TGCTGTGGTA ACTGCTGACG GGAAGCTTGC TTAAACCAG	480
	ATTAGCCAGA TCTCCATGAA GTCGCCACAG CTGATTTGG TGAATATGGC CAGCTCCC	540
15	GAGTGTACAG CTGCAGCTAT CAAGGCTATA AGAGAAAGTG GAATGAATCT GAACCCAGAA	600
	GTGGAAGGGA CGCTAATTG GGTACCCATT CCCCAGTAA CCAGAGAGCA CAGAGAAATG	660
20	CTGGTGAAAC TGCCAAACA GAACACCAAC AAGGCCAAAG ACTCTTACG GAAGGTTCGC	720
	ACCAACTCAA TGAACAAGCT GAAGAAATCC AAGGATACAG TCTCAGAGGA CACCATTAGG	780
	CTAATAGAGA AACAGATCAG CCAAATGGCC GATGACACAG TGGCAGAACT GGACAGGCAT	840
25	CTGGCAGTGA AGACCAAAGA ACTCCCTTGA TGAAAGTCCA CTGGGGCCAG CAATACTCCA	900
	GAGCCCAGTT TCTGCTGGAT CCCATGGGTG GCACATTGGG ACTTCTCTCC CTCCCCATC	960
30	TACACAGAAG ACTGTACCCA TGCTGACAGA AGCCTGTCCT TGTAGGCC AGCCTTCCAG	1020
	GGGAACACTC AGACATGTC ATTCTCTTCC TGCTCTGCT CTGGGCCGGT GGGTGGCTCT	1080
	CAGAAAWAC TTGCTGCTGG CAAAAGGCCT GTACTCAGGC ATTTGCTTIG ACTTGATGTT	1140
35	GCCAAGGGAC TGAGGCCATT GGCAGGCTTA GTACCACCTG CTCCCTCATCT TAGGAGCTC	1200
	CTTTCTAAAT AATTAGGCTC TGTCTCCATT TTAAACTCT GATATTGCC TTCACCTGTG	1260
40	ACTGGACACT TTACTAGAGG CCCATTTCA CTAAACAATA AAATCTAAAT AAATTGGAAG	1320
	GAATAACAAC CACAAAGGAA AGAATAGAGT TGGCTCTGGAT TGATGATCAC TGAGGAATCTG	1380
	TATGTGAGGC ACCCATAACA GTAGTTTGC CTGTGAGTCG TCTTCACACA TGCTGTTTC	1440
45	TCTGCCTGGC TCTCTCTTCC CCTCCTTACC TGGCCAGTCC TGTCTTATCAT CAGGCCCTGT	1500
	CTTGGATATC ACGTCTCTG GGAAGTCITC TTTTCCCTC TAACCTAGGA CCCTCATTAC	1560
50	CGGCTCTCAT AGCACAGTCT ACTGCTTTGT ACAGAATTCTA AGTATTCTTG TTGCACTTAA	1620
	TTAGGCCTGTA TATCCTCAGA ACTTTGCTGTA ATGCCTGGAG CATACTAGGC AGTCATATGT	1680
	TGTATCGTGA ATAAATTGCA CATACTAGCT ACCCAGCAA TGCTGACTTC TTTTCTTCT	1740
55	AGTCTTAACA CTCCCTTCT AATNCATTTC CACTMTGTA MTGTTCTCAA CATTACTTGG	1800
	TAGTGTACAAA CTTT	1814

(2) INFORMATION FOR SEQ ID NO: 74:

5	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 4712 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:	
	CATGGTACGC CTGCAGGTAC CGGTCCGGAA TTCCCCGGTC GACCCACGCG TCCGCCCAYG	60
15	CGTCGGCGG CTCCGAGCCA GGGCTATTG CAAAGCCAGG GTGCCCTACC GGACGGAGAG	120
	GGGAGAGCCC TGAGCAGAGT GAGCAACATC GCAGCCAAGG CGGAGGCCGA AGAGGGGCC	180
	CAGGCACCAA TCTCCGGTT GCCTCAGCCC CGGAGGCCGC CCAGAGCGCT TCTTGTCCCA	240
20	GCAGAGCCAC TCTGCMTCG CCTGCCTCTC AGTGTMTCCA ACTTTGGCTT GGAAGAAAAA	300
	CTTCCCGCGC CGCGGCAGAA CTGCAGCGCC TCCCTCTTAGT GACTCCGGGA GCTTCGGCTG	360
25	TAGGCCKGCTM TGCGCGCCCT TCCAACGAAT AATAGAAATT GTTAATTITA ACAATCCAGA	420
	GCAGGCCAAC GAGGCTKTGC TCTCCCGACC CGAACTAAAG CTCCCTCGCT CCGTGCGCTG	480
	CTACGAGCGG TGTCTCCTGG GGCTCCAATG CAGCGAGCTG TGCCCGAGGG GTTOGGAAGG	540
30	CGCAAGCTGG CGAGCGACAT GGGGAACGCG GAGCGGGCTC CGGGGTCTCG GAGCTTTGGG	600
	CCCGTACCCA CGCTGCTGCT GCTCSCCCGC GCGCTACTGS CCGTGTGCGA CGCACTCGGG	660
35	CGCCCCCTCCG AGGAGGACGA GGAGCTAGTG GTGCCGGAGC TGGAGCGCGC CCCGGGACAC	720
	GGGACCACGC GCCTCCGCCT GCACGCCCTT GACCAGCAGC TGGATCTGGA GCTGCGGCC	780
	GACAGCAGCT TTTTGGCGCC CGGCITCAGG CTCCAGAACG TGGGGCGCAA ATCCGGGTCC	840
40	GAGACGCCGC TTCCGGAAAC CGACCTGGCG CACTGCTCT ACTCCGGCAC CGTGAATGGC	900
	GATCCAGCT CGGCTGCCGC CCTCAGCCTC TGCGAGGGCG TGCGCGGCC CTTCTACCTG	960
45	CTGGGGGAGG CGTATTTCAT CCAGCCGCTG CCCGCCGCCA GCGAGCGCCT CKCCACCGCC	1020
	GCCCCAGGGG AGAAGCCGCC GGCACCACTA CAGITCCACC TCCCTGGCGC GAATCGGCAG	1080
	GGCGACGTAG CGGGCACGTG CGGGGTCTGTG GACGACGAGC CCCGGCGAC TGGGAAACCG	1140
50	GAGACCGAAG ACGAGGACGA AGGGACTGAG GGCGAGGACG AAGGGCCTCA GTGGTCGCCG	1200
	CAGGACCCGG CACTGCAAGG CGTAGGACAG CCCACAGGAA CTGGAAGCAT AAGAAAGAAG	1260
55	CGATTTGTGT CCAGTCACCG CTATGTGGAA ACCATGCTTG TGGCAGACCA GTCGATGGCA	1320
	GAATTCCACG GCAGTGGCT AAAGCATTAC CTTCTCACGT TGTTTCTGGT GGCAGGCCAGA	1380
	TTGTWCAAAC ACCCCAGSAT TCGTAATTCA GTTACGCTGG TGGTGGTGAA GATCTTGGTC	1440
60	ATCCACGATG AACAGAAGGG GCCGGAAGTG ACCTCCAATG CTGCCCTCAC TCTGCGGAAC	1500

	TTTTGCAACT GGCAAGAGCA GCACAACCCA CCCAGTGACC GGGATGCAGA GCACATGAC	1560
5	ACAGCAATTG TTTCACCAAG ACAGGACTTG TGTGGGTCCC AGACATGTGA TACTCTGGG	1620
	ATGGCTGATG TTGGAACGTG GTGTGATCCG AGCAGAAGCT GCTCCGTCA AGAAGATGAT	1680
	GGTTTACAAG CTGCCTTCAC CACAGCCAT GAATTAGGCC ACGTGTTAA CATGCCACAT	1740
10	GATGATGCAA AGCAGTGTGC CAGCCTTAAT GGTGTGAACC AGGATTCCCA CATGATGGCG	1800
	TCAATGCTT CCAACCTGGA CCACAGCCAG CCTTGGTCCTC CTTGCAGTGC CTACATGATT	1860
15	ACATCATTTC TCGATAATGG TCATGGGAA TGTGTGATGG ACAAGCCTCA GAATCCCATA	1920
	CAGCTCCAG GCGATCTCCC TGGCACCTCG TACGATGCCA ACCGGCAGTG CCAGTTTACA	1980
	TTTGGGAGG ACTCCAAACA CTGCCCTGAT GCAGCCAGCA CATGTAGCAC CTTGTGGTGT	2040
20	ACCGGCACCT CTGGTGGGT GCTGGTGTGT CAAACCAAAC ACTTCCCGTG GGCGGATGGC	2100
	ACCAAGCTGTG GAGAAGGGAA ATGGTGTATC AACGGCAAGT GTGTGMAAA AACCGACAGA	2160
25	AACCAATTTC ATACCCCTTT TCATGGAAGC TGGGAATGT GGGGGCCCTTG GGGAGACTGT	2220
	TCGAGAACGT GCGGTGGAGG AGTCCAGTAC ACCATGAGGG AATGTGACAA CCCAGTCCCA	2280
	AAGAATGGAG GGAAGTACTG TGAAGGCCAA CGAGTGCCTG ACAGATCCG TAACCTTGAG	2340
30	GACTGTCCAG ACAATAATGG AAAAACCTTT AGAGAGGAAC AATGTGAAGC ACACAAACGAG	2400
	TTTCAAAAG CTTCTTTGG GAGTGGCCT GCGGTGGAAT GGATTCCCA GTACGCTGGC	2460
35	GTCTCACCAA AGGACAGGTG CAAGCTCATC TGCCAAGCCA AAGGCATTGG CTACTCTTC	2520
	GTGGTGCAGC CCAAGGTGT AGATGGTACT CCATGTAGCC CAGATTCCAC CTCTGTCTGT	2580
	GTGCAAGGAC AGTGTGTAAA AGCTGGTTGT GATCGCATCA TAGACTCCAA AAAGAAGTT	2640
40	GATAATGTG GTGTTGCGG GGGAAATGGA TCTACTTGTAA AAAAATATC AGGATCAGTT	2700
	ACTAGTGCCTT AACCTGGATA TCATGATATC ATCACAAATTG CAACTGGAGC CACCAACATC	2760
45	GAAGTGAACAC AGCGGAACCA GAGGGGATCC AGGAACAAATG GCAGCTTTCT TGCCATCAA	2820
	GCTGCTGATG GCACATATAT TCTTAATGGT GACTACACTT TGTCCACCTT AGAGCAAGAC	2880
	ATTATGTACA AACGTGTGTG CTTGAGGTAC AGCGGCTCCT CTGGGGCATT GGAAAGAATT	2940
50	CGCAGCTTAA GCCCTCTCAA AGAGCCCTTG ACCATCCAGG TTCTTACTGT GGGCAATGCC	3000
	CTTCGACCTA AAATTAATA CACCTACTTC GTAAAGAAGA AGAAGGAATC TTCAATGCT	3060
55	ATCCCCACTT TTTCAGCATG GGTCAATTGAA GAGTGGGGCG AATGTCTAA GTCATGTGAA	3120
	TTGGGTTGGC AGAGAAGACT GGTAGAATGC CGAGACATTA ATGGACAGCC TGCTTCCGAG	3180
	TGTGCAAAGG AAGTGAAGCC AGCCAGCACC AGACCTTGTG CAGACCATCC CTGGCCCCAG	3240
60	TGGCAGCTGG GGGAGTGGTC ATCATGTTCT AAGACCTGTG GGAAGGGTA CAAAAAAAGA	3300

	AGCTTGAAGT GTCTGTCCCA TGATGGAGGG GTGTTATCTC ATGAGAGCTG TGATCCTTA	3360
5	AAGAAACCTA AACATTTCAT AGACTTTGC ACAATGGCAG AATGCAGTTA AGTGGTTAA	3420
	GTGGTGTAG CTITGAGGGC AAGGCAAAGT GAGGAAGGGC TGGTGCAGGG AAAGCAAGAA	3480
	GGCTGGAGGG ATCCAGCGTA TCTTGCCAGT AACCAAGTGAG GTGTATCAGT AAGGTGGAT	3540
10	TATGGGGTA GATAGAAAAG GAGTTGAATC ATCAGAGTAA ACTGCCAGTT GCAAATTGTA	3600
	TAGGATAGTT AGTGAGGATT ATTAACCTCT GAGCAGTGAT ATAGCATAAT AAAGCCCCGG	3660
15	GCATTATTAT TATTATTCTCT TTGTTACAT CTATTACAAG TTAGAAAAAA ACAAAAGCAAT	3720
	TGTCAAAAAAA AGTTAGAACT ATTACAACCC CTGTTTCTG GTACTTATCA AATACTTAGT	3780
	ATCATGGGGG TTGGAAATG AAAAGTAGGA GAAAAGTGAG ATTTTACTAA GACCTGTTT	3840
20	ACTTTACCTC ACTAACAAATG GGGGGAGAAA GGACTACAAA TAGGATCTT GACCAGCACT	3900
	GTTTATGGCT GCTATGGTTT CAGAGAATGT TTATACATTA TTCTACCGA GAATTAACAC	3960
25	TTCAAGATTGT TCAACATGAG AGAAAGGCTC AGCAACGTGA AATAACCCAA ATGGCTTCCT	4020
	CTTTCTTTT TTGGACCATC TCAGTCTTTA TTGTTGTAAT TCATTTGAG GAAAAAACAA	4080
	CTCCAATGAT TTATTCAAGT GCATTTAAAGT CTACAATGGA AAAAAGCAG TGAAGCATT	4140
30	GATGCTGGTA AAAGCTAGAG GAGACACAAT GAGCTTAGTA CCTCCAACCT CCTTTCTTTC	4200
	CTACCATGTA ACCCTGCTTT GGGAAATATGG ATGTAAGAA GTAACTTGTG TCTCATGAAA	4260
35	ATCACTACAA TCACACAAGG AGGATGAAAC GCCGGAACAA AAATGAGGTG TGTAGAACAG	4320
	GGTCCCACAG GTTGGGGAC ATTGAGATCA CTTGTCCTGT GGTGGGGAGG CTGCTGAGGG	4380
	GTAGCAGGTG CATCTCCAGC AGCTGGTCCA ACAGTCGTAT CCTGGTGAAT GTCTGTTCAG	4440
40	CTCTCTGTG AGAATATGAT TTTTCCATA TGTATATAGT AAAATATGTT ACTATAAATT	4500
	ACATGTACTT TATAAGTATT GTTTGGGTG TTCCCTCCAA GAAGGACTAT AGTTAGTAAT	4560
45	AAATGCCAT AATAACATAT TTATTTTAT ACATTTATTT CTAATGAAAA AAACCTTTAA	4620
	ATTATATCGC TTTGTGGAA GTGCATATAA AATAGAGTAT TTATACAATA TATGTTACTA	4680
	GAAATAAAAG AACACTTTTG GAAAAAAA AA	4712
50		

(2) INFORMATION FOR SEQ ID NO: 75:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1885 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

	ATGCCARGAA GACTGATGGA GCAGGCTTGC AATATTAAAG TNCCAACCAA GAAGCTGAAG	60
5	AAATWTGAGA AAGAATATCC AGACAATGCG AGAGAGTCAG CTGCAACAGG AAGACCCAAT	120
	GGATAGATAC AAGTTTGTAT ATTTGTAGGT AACTCCAGCT GTTGCATTTA TACTGGGAAT	180
10	CTTCATAAGA AGCTGAGAGA AAGAGAGGGG AAAAGAAG TGGCITTCTA CTTTCAAAAA	240
	TGAAACAAA AGGAAAAATG GCAAAGTACT GTTTAGCTG TGCAATGTCAT ATCCACAAAG	300
	ACTTTTAGCA GGTGAACGT TCCAAGACTG ACACAAGGAT GTTCAAAC TGCCTCTGTC	360
15	TGTAGAAAAT GTAAAAATA CCAACTCACT TGGAGGAAA AATAAAAATC ACAAAGGTAT	420
	ATTGAGCACA GTAGTGGTGT TTGTTGCAAC ATTTATTCAC ACAAATGAAT TTATGAACAA	480
20	CAGTGATATT TGACTTAAAG TATGAAGTTT CAGAATCAA ATAATTCAT TTTAATACGT	540
	TCNGTTAATT GTGAATCTCT TCMATGGTAA TTAGCAACAC TGTTCCCAGG ATGCAAAGTT	600
	GGGAAACACT TATTTCAAC TTATTTTTT CCAAGTAAA TATTATCTCT CPTCAACATG	660
25	CTTTAACCTT TCAGACTCAC ACAGATACTG WACAGCTCCC TTCTCCCTCC ATATCAATAC	720
	ACTAAGATAA AAGAATACTG TATTTTCAGC ACTGAGCAGC AGTGCACAAA TCTCCTGCCA	780
30	AGAAATGGAC TGIGTGGCAT TATTAATTAA ATCACCCACA TTGGGATGAC TTCCACTTTT	840
	GTAACTAGAG TTATCTTTAT GTGGTCAGAG CTGGACATAG GCAGCATAGT CACACAGAAC	900
	ATCTTATCTC TGTGCKGAA TKGAATAGCA TGGGATGTC GTAGAGAAC ATGGGGGAG	960
35	TATGTAGTTT TKGAGTCAG ACAGACCKGA ACTCAAATCT TGTCATTTT TTAGAGCACA	1020
	GGATTTCGAY TCCAAATTGA GGGTTTTAAT CCCCCATGCCA CCATTGAGCA TCTTCGACTA	1080
40	GTTATTGAAC CTYTTCCCTCA TSKATAAAAG ATATAGTGT TCTGATTCT TGATGGATTG	1140
	TTACAAGGAT GAGGGATGCT GTATGTTAAG GACTCAGCTC ATAGTGTGTT TCAATAAAATG	1200
	GCTGTTATT TATGAAGCCT ACTACTACAG ATTATGCAAT TATTACTAGA ATAATGCCAC	1260
45	CTTATGTGGG TCTTCCCTTC TAGTCCCTTA TTGATTGTC TTATTTCTCT CAAGTATTGC	1320
	CAACCAATAA TCTCCCTTG CTTATAGAAG TGGTCAAGA TCTGATTATA AAATCCCACA	1380
	TACTCTATA GCAGATAACT ATTAACAGAT AATGTTGTA CTAATTCAC CACCAACATT	1440
50	CCCCCTCAAT AAAACCAAGCT TTAAATGTAATCACATAGC ATACTGCTTT AGAAAGGCTT	1500
	GAAGGTAGTA ATTATAAACT ATTATTAAGC ATCCAAAATG AAGGTCTCCT TTTGCTAATA	1560
55	TCATTTCAGAT TTCTTATTA CTACAATTAT TATGAATAA TTCTGTGAAG AGTGCCTTAA	1620
	AATAAGAGAG AAATGGRAGA CCAAACCTGT ACATTTAAA TCAGGCTGGA ATTGAACCTG	1680
60	TTATGTGTC TAAATCCCTT TTGTTGCCA AAGCAGGTAT GTATACATTA ATAGTAAGAT	1740

	GTACATTATT TTAAAGTAC TTATMACATG TAAGATTATC AATATGTATA GTTTTATTG	1800
	AGAGATCAA GTAGGATTAA ACITCTTGT TTGAAACAG GCATTACTTT TTAAAAAAA	1860
5	AAAAAAA AAAA AAAA AAAA	1885

10 (2) INFORMATION FOR SEQ ID NO: 76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 890 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

20	TICAAACTAG CAAAAATGT ATGAAACTAT GAAGCTCGAT GCGTGTRATC ATCAGCAGAG	60
	GGCGACGCTG CAGGCAGGGC CAAAGCTTCT GACCCCTGGCC CCCAGGGAGG AACCCAGAGG	120
25	CCAGTCACGG AGGCCCCAGCG AGCTCACCGG CAGGCCAGGGC CACAGCAGTC GCGACCCCTCA	180
	GGGAGAACAG GCACTACCCA GGGCTGGATG CGTAACGGGC CCCCCGGCCA CACCCCACCG	240
	CCCATCAGAG CGCGAGCTCC TGAGAACGCA TCCGGATGCN AGGCCAAGT CAGCCATGGC	300
30	ACAAACATTT GTGCATCAAG GTCCCTGTG TCTGCAACAA CTCACCACAA ACAGAAGGGT	360
	GGAAACCTCC ATGTCATCGG ACGGCCACGG SCAGAACCCA ACGCCATCTC CCTGGGCTGA	420
35	TGTCCTGTCA AGCAGGGCTG ATGCCGTAGC TTTTCCGGCT TCTGGAARCT GCCACAGCCC	480
	CTGGCTCATG GSACCATCCT CACATCCTCT GAATCCACAT TCTCCTCTGA ATCTCCCGCC	540
	TCCCTCTTTC CACTGTAAGG ACCCTGTGAT GACACTGCAC CCTCAGACCC TGGTAACCCA	600
40	GGGTCACTTT TCCACCTCAG GGCCTCTGAC TTAAGCCTGC CTGGAGGGTC CCTGTGGTCA	660
	CATTCAATGGG TTCCAGGCTT CAGACACGGC CACTTTGTGG GATCATTACT CTGCCTACCA	720
45	CACCATGTGG CCCTGTGTGT GTTTTCAGGG GGCATTTGCG CYTATATGCA AATAATACAT	780
	ATATGAATAA ACGTGTGAAT GGTGGTCACG TAGGAGARGG CATCTGTATG GGGCCACACC	840
	TGTAAAAAAA AAAA	890
50		

(2) INFORMATION FOR SEQ ID NO: 77:

55	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1657 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

	AGAACGGCCT TCCCCACATC TTCCAGCACC TCGCGCCTG AATCCGTCCC ACCCAGGCC	60
5	AGACGCAGGC TTCTTCTCGG GTCTTGGTCC TGCATCTCT CTCTCCAGA GCCTCCGTTA	120
	GGGGTGGGAA AGGACTTTGC CATAGGTGCG TGAGGCCACC ATCTGCTCTC TTACTGGCCA	180
10	AGGGCGTAAA AAGATAGTCY TCCCATTAGC TAGAGAGCAA ACCCCAGAAA GCCTATTGGC	240
	TGCGCCGTCC GCGGGCCTTG GTCCGNTTTG AAGGCCGCT GCGGCTGCGA GAGGAGGGCG	300
	GGCGGGAGGC TAGCTGTTGT CGTGGTTGCT CGGAGGCACG TGTGCACTCC CGGAAGCGGC	360
15	GAGGGAAAC TGCTCCGCGC GCGCCGCGGG AGGAGGAACC GCGCGTCCT TTAGGGTCGG	420
	GGCCCGGGCG GGCATGGATT CAATGCCTGA GCCCGCGTCC CGCTGCTCTC TGCTTCCTCC	480
20	CTTGCTGCTG CTGCTGCTGC TGCTGCTGCC GGCCCCGGAG CTGGGCCCCGA GCCAGGCCGG	540
	AGCTGAGGAG AACGACTGGG TTGCGCTGCC CAGCAAATGC GAAGGGACTT GCGGTTAATC	600
	GAAGTCACTG AGAACCAATTG GCAAGAGGCT CCTGGATTAT AGCCTGCACA AGGAGAGGAC	660
25	CGGCAGCAAT CGATTTGCCA AGGGCATGTC AGAGACCTTT GAGACATTAC ACAACCTGGT	720
	ACACAAAGGG GTCAAGGTGG TGATGGACAT CCCCTATGAG CTGTGGAACG AGACTTCTGC	780
	AGAGGTGGCT GACCTCAAGA ACCAGTGTGA TGTGCTGCTG GAAGAGTTTG AGGAGGTGAT	840
30	CGAGGACTGG TACAGRAACC ACCAGGAGGA AGACCTGACT GAATTCCCTC GCGCCAACCA	900
	CGTGTGAAG GGAAAAGACA CCAGTTGCC GGCAGAGCAG TGGTCCGGCA AGAAGGGAGA	960
35	CACAGCTGCC CTGGGAGGG AAGAAGTCCAA GAAGAAGAGC AKCAGGGCCA AGGCAGCAGG	1020
	CGGCAGGAGT AGCAGCAGCA AACAAAGGAA GGAGCTGGT GGCTTGAGG GAGACCCCAG	1080
40	CCCCGAGGAG GATGAGGGCA TCCAGAAGGC ATCCCCCTCTC ACACACAGCC CCCCTGATGA	1140
	GCTCTGAGCC CACCCACCAT CCTCTGTCCT GAGACCCCTG ATTTTGAAGC TGAGGAGTC	1200
	GGGGCATGGC TCTGGCAGGC CGGGATGGCC CGCAGCCTT CAGCCCCCTC TTGCTTGGC	1260
45	TGTGCCCCCTCTCTC TCTGCAAGG AAAGACACAA GCCCCAGGAA GAACTCAGAG CGTCATGGG	1320
	TAGCCCACGC CGTCCTTTC CCTCCCCAAG TGTTTCTCTC CTGACCCAGG GTTCAGGCAG	1380
	GCCTTGTGGT TTCAAGGACTG CAAGGACTCC AGTGTGAACG CAGGAGGGC AGGTGTCAGA	1440
50	ACTGGGCACC AGGACTGGAG CCCCTCCGG AGACCAAACG CACCATCCCT CAGTCCTCCC	1500
	CAACAGGGTA CTAGGACTGC AGCCCCCTGT AGCTCTCTC TGCTTACCCC TCCGTGGAC	1560
55	ACCTTGCAGT CTGCTGGCC CTTCCCCAGAG CCCAAAGAGT AAAATGTTG TGGTCTGAW	1620
	AAAAAAAAA AAAAAAAAAA CCCGGGGGGG GGCCCGT	1657

(2) INFORMATION FOR SEQ ID NO: 78:

	(i) SEQUENCE CHARACTERISTICS:	
5	(A) LENGTH: 2015 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
10	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:	
	GGCCGGGCTG AGAGAAGAGC TTGCGGGT TCGGGTGTAT GGCCCCGACT GAAGGGCTGG	60
15	AGGCGGTGTA TGGCGCTGTT CTTGCTGTGCG CTCCCGACAC CTCCGTCCGC TTCTGGTCAT	120
	GAGAGGAGAC AGAGGCTGAG AGCAAAGACA TCTGGGTCAAG AGAAAAAGTA TTAAAGGCC	180
	ATGCAAGCCA ATCGTAGCCA ACTGCACAGT CCTCCAGGAA CTGGAAGCAG TGAGGATGCC	240
20	TCAACCCCTC AGTGTGTCCA CACAAGAATG ACAGGAGAGG GTTCTTGCCC TCATTCTGGA	300
	GATGTTCATATAA CTCCATACCT AAAGAATGTG CAGAAAATGC AAGCTCCAGA	360
25	ATATATAAGGT CAGGTGTCCA TAGCTGTGCC CATGGATGTG TACACAGTCG CTTACGGGT	420
	CACTCCCACA GTGAAGCAAG GCTGACTGAT GATACTGCCG CAGAATCTGG AGATCATGGT	480
	AGTAGCTCCT TCTCAGAATT CGCTATCTC TTCAAGTGGC TCCAAAAAAG TCTTCCATAT	540
30	ATTTGATTC TGACCGTCAA ACTTGTTATG CAGCATATAA CAGGAATTTC TCTTGGATT	600
	GGGCTGCTAA CAACTTTAT GATGCAAAC AAAAGCATTG TAAATCAGGT TTTTCTAAGA	660
35	GAAAGGTCTT CAAAGATTCA GTGTGCTTGG TTACTGGTAT TCTTAGCAGG ATCTTCTGTT	720
	CTTTTATATT ACACTTTCA TTCTCAGTCA CTTTATTACA GCTTAATTTT TTTAAATCCT	780
	ACTTTGGACC ATTTGAGCTT CTGGGAAGTA TTTKGGATTG TTGGAAATNAC AGACTTCATT	840
40	CTGAAATTCT TTTTCATGGG CTTAAATGC CTTATTTAT TGGTGCCTTC TTTCATCATG	900
	CCTTTAAAT CTAAGGGTTA CTGGTATATG CTTTTAGAAG AATTGIGTCA ATACTACCGA	960
45	ACTTTTGTTC CCATACCAAGT TTGGTTTCGC TACCTTATAA GCTATGGGA RTTTGGTMAC	1020
	GTAACTAGAT GGARTCTTGG GATACTGCTG GCTTTACTCT ACCTCATATT AAAACTTTTG	1080
	GAATTTTTG GGCATCTGAG AACTTTAGA CAGGTTTAC GAATATTTT TACACMACCM	1140
50	AGTTATGGAG TGGCTGCCAG CAAGAGACAG TGTTCAAGATG TGGATGATAT TTGTTCAATA	1200
	TGTCAAGCTG AATTTCAAGAA GCCAATTCTT CTCATTGTC AGCATATATT TTGIGAAGAG	1260
55	TGCATGACCT TATGGTTAA CAGAGAGAA ACATGTCCAC TCTGCAGAAC TGTGATTCA	1320
	GACCATATAA ACAAAATGGAA GGATGGAGCC ACTTCATCAC ACCTTCAAAT ATATTAAGTT	1380
	GTATAAACTA TCAAGGCCAC AAAATACTAA TGTCAATTGG TCATAATGAC TACTGATAAG	1440
60	GCATCAGAAT GGATTTTCAG GGCTACCAGA AAAATGTTTC CAGATGGTT TAGAATGTAG	1500

	GA CTTATGAT CCAATTCA CC AAAGATTAA ATGAAACCAC CCTGTGTTT AAAATATA	1560
5	TAATGTC AA CCTAATGTAT ATGCAACATT TATTCTATTC TAATTATTTG ACAGGTA ACT	1620
	GCAGTGT AA ATTGTAAATG TGTTTCTTT ATGTTACCAA AACAGCAATT TGAAATTAGA	1680
	ACTAGTGGTT TTAGAGAACT CAGGTATTCT TTCC CTGACAT TGTTTCAGA ATAAAGAATA	1740
10	TTTT TCATAA TATTTAAGA TACATACTAT CTAAAAGTAG AATT TGTTC AGCAATTGACT	1800
	TTTATAATT C C AATCTTAA AATTCTTAA ATTTCTAA AATTCTAAAT GT GTGTATT ATCAGTAACA TTTTCTAAGT GAAGATTAA TTACTGAGGA	1860
15	TGATACATTA TAGTATTGTA TTATTCTCTG TAGTAAGATT AGTAATAAGT GAAAATAAAT	1920
	GATT AAATT CAAAAAAA AAAAANTNA CTCGA	1980
20		2015

(2) INFORMATION FOR SEQ ID NO: 79:

25	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1213 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:	
	AGCCTAGTTA CAGATTGCAC TGC GT GTACAGAC TGTTCCACAC CCAGAAGACG TCAGGTGACT	60
35	TCAGTCCTGC TGCAGTTGTG CAGCAGAGGA GACTGCAGAC TT CGGTTGAG GAAACGGGTA	120
	TTTCATGTCT CAGGGAGTAG GTTTGTGCAG TTACAGCTTT TCTGTGGTA TGCATAATT A	180
40	ATAATTGGAG CTGCAAASCA GATCGTGACA AGAGATGGAC CGTCAGAAGA AAAATGGAA GGACAAAGTT GTTGACCTCC TGTACTGGAG AGACATTAAAG AAGACTGGAG TGGTGT TGG	240
	TGCCAGCCTA TT CCTGCTGC TTTCATTGAC AGTATT CAGC ATTGTGAGCG TAACAGCCTA	300
45	CATTGCTTG GCCCTGCTCT CTGTGACCAT CAGCTT TAGG ATATA CAAGG AGTCA AGCC 420	360
	AGCTATCCAG AAATCAGATG AAGGCCACCC ATT CAGGGCA TATCTGGAAAT CTGAAGTTGC	480
50	TATATCTGAG GAGTTGGTTC AGAAGTACAG TAA TCTGCT CTTGGTCATG TGA ACTGCAC GATAAAAGGAA CTCAGGCGCC TCTTCTT AGT TGATGATT A 540	600
	AGTGGTGTGAG TG GGTATT TA CCTATGTG GG TG CCTTGTTT AATGGTCTGA CACTACTGAT	660
55	TTTGGCTCTC ATT TCACTCT TCAGTGT CC TGTTATT AT GAA CGGCATC AGGCACAGAT AGATCATTAT CTAGGACTTG CAAATAAGAA TG TAAAGAT CCTATGGCTA AAATCCAAGC	720
	AAAAATCCCT GGATTGAAGC GCAAAGCTGA ATGAAAACGC CCAAAATAAT TAGTAGGAGT	780
60		840

	TCATCTTTAA AGGGATATT CATTGATTA TACGGGGAG GGTCAAGGAA GAACGAACCT	900
	TGACGTTGCA GTTCAGTTTC ACAGATCGTT GTTAGATCTT TATTTTTAGC CATOCACTGT	960
5	TGTGAGGAAA AATTACCTGT CTTGACTGCC ATGTGTTCAT CATCTTAAGT ATTGTAAGCT	1020
	GCTATGTATG GATTTAAACC GTAATCATAT CTTTTTCTA TCTGAGGCAC TGGTGAATA	1080
10	AAAAACCTGT ATATTTACT TTGTTGCAGA TAGTCTTGCC GCATCTTGGC AAGTTGCAGA	1140
	GATGGTGGAG CTAGAAAAAA AAAAAAAA ANCTYGAGAC TAGCGGCCACG AGGGGGGCC	1200
	CGTACCCAN ACG	1213

15

(2) INFORMATION FOR SEQ ID NO: 80:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1391 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

	GCAGAGGCCG ACTGCTGAAG GTGGTTGCGC TCGACATGGC GGTTACCCCTG AGTCTCTTGC	60
30	TGGCGGGCG CGTTTGGCGC CCGTCACTCG CTGTGGGTTG GCGACCCGGG CGGTGGCGGG	120
	CCCAGGCCCT ATTGGCCGGG AGCCGGACCC CGATTCCGAC TGGGAGCCGG AGGAACGGGA	180
35	GCTGCAGGAG GTGGAGAGCA CCCTGAAACG ACAGAAACAA GCAATCCGAT TCCAGAAAAT	240
	TCGGAGGCAA ATGGAGGCCG CTGGTGCCCC GCCCAGGACC CTGACGTGGG AAGCCATGGA	300
	GCAGATAACGG TATTTACATG AGGAATTTCAG AGAGTCCTGG TCAGTCCCA GGTTGGCTGA	360
40	AGGCTTGTAT GTCAGCACTG ATGTGATCCG AAGAGTTTA AAAAGCAAGT TTTTACCCAC	420
	ATTGGAGCAG AAGCTGAAGC AGGATCAAAA AGTCCTTAAG AAAGCTGGC TTGCCCACTC	480
45	GCTGCAGCAC CTCCGGGCT CTGGAAATAC CTCAAAGCTG CTCCCTGCAG GCCACTCTGT	540
	ATCAGGCTCT TTGCTTATGC CAGGGCATGA AGCCTCATCT AAAGACCCAA ATCACAGCAC	600
	AGCTTGAAA GTGATAGAGT CAGACACTCA CAGGACAAAT ACACCAAGGA GAAGGAAGGG	660
50	AAGAAATAAA GAAATCCAGG ACCTGGAGGA GAGCTTGTG CCTGTTGCTG CACCCCTAGG	720
	TCATCCAAGA GAGCTGCAGA AGTACTCCAG TGATTCTGAG AGCCCCAGAG GAACTGGCAG	780
55	TGGTGCCTG CCAAGTGGTC AGAACGCTGGA GGAGTTGAAG GCAGAGGAGC CAGATAACTT	840
	CAGCAGCAA GTAGTGCAGA GGGCCCGAGA GTCTTTGAC AGCAACGGGA ACTTCCTGTA	900
	CAGAATTGAGA GTCGGGCTT GGCTTATGGA GATGCCTCGT GAAACACAGC TGGGCAAGTA	960
60	TTAATGTATA TGGAACAGCC TGGATTCTG CATAATGGATA AGCCACCTTG GAATAGGAAG	1020

	AGGTGTTGAG CCTGGACTGT GGGAGGAAAG AGCTGCGTGG ATAGATTCAA ACTTCCGTG	1080
5	GTAGTGCCTCC CAGTCTGACC TCTGTAGACC TTCAGTACTC ACTCTTCTTG CTTAGGCTCT	1140
	CTGTGTTGAAAGCCATCC CGTGTGTCAT GTGTTGTTAC AATTTCTGT GATACTTGCA	1200
	ATTTATGTTT GAGAAGAAGT GAAAAGTTG CCTTCTGACC TCATTTCTT CTTGATCAGT	1260
10	GAACACTAAC ATTTGGGGA CAACTTAGTC AATTGGTTT CCTTACAACA AAATAAAGTA	1320
	AAATGTAGCA AAAAAAAA AAAAAAAACN CCGGGGGGGC CCGTCCCATT GCCCAAAAGG	1380
15	GGCCGAATA A	1391

20 (2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1008 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

30	TGACATCGCC CTCATGAAGC TGCAGTCCCC ACTCACITTC TCAGGCACAG TCAGGCCAT	60
	CTGCTGCCCTTCTTGATG AGGAGCTCAC TCCAGCCACC CCACCTCTGGA TCATTGGATG	120
	GGGCTTTACG AAGCAGAATG GAGGGAAAGAT GTCTGACATA CTGCTGCCAGG CGTCAGTCCA	180
35	GGTCATTGAC ACCACACGGT GMAATGCAGA CGATGCGTAC CAGGGGAAG TCACCGAGAA	240
	GATGATGTGT GCAGGCATCC CGGAACGGGG TGTGGACACC TGCCAGGTG ACAGTGGTGG	300
40	GCCCCCTGATG TACCAATCTG ACCAGTGGCA TGTGGTGGC ATCGTTAGCT GGGGCTATGG	360
	CTGCCGGGGC CGGAGCACCC CAGGAGTATA CACCAAGGTG TCACGCTATC TCAACTGGAT	420
	CTACAATGTC TGGAAAGGCTG AGCTGTAATG CTGCTGCCCTTTCAGTGC TGGGAGCCGC	480
45	TTCCCTTCTG CCCTGCCAC CTGGGGATYC CCCAAAGTCA GACACAGAGC AAGAGTCCCC	540
	TTGGGTPACAM CCCTYTGCCAC ACAGCCTCAG CATTTCCTGG AGCAGCAAAG GGCCTCAATT	600
50	CCTATAAGAG ACCCTCGCAG CCCAGAGGCG CCCAGAGGAA GTCAGCGGCC CTAGCTCGGC	660
	CACACTTGGT GCTCCCAGCA TCCCAGGGAG AGACACAGCC CACTGAACAA GGTCTCAGGG	720
	GTATTGCTAA GCCAAGAAGG AACTTTCCCA CACTACTGAA TGGAAGCAGG CTGCTTGTAA	780
55	AAAGCCCAGA TCACGTGGG CTGGAGAGGA GAAGGAAAGG GTCTGCCA GGCCTGTCCG	840
	TCCTCACCCA TCCCCAAGCC TACTAGAGCA AGAAACCAGT TGTAATATAA AATGCACTGC	900
60	CCTACTGTG GTATGACTAC CGTTACCTAC TGTGTCATT GTTATTACAG CTATGCCAC	960

TAATATTAAC GAGCTGTGTA ACATCAAAAA AAAAAAAA AAACTCGA	1008
----------------------------------------------------	------

5

(2) INFORMATION FOR SEQ ID NO: 82:

10	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1261 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
15	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:	
	GTTCATTCCTT CCCGTCTCCT TCATAGAATA CTACTTTTC CTTTTGTCCTC CTGGCCATT	60
	GGAATTCAAGG CTCACCAATTG GCGGAGGCAA GCCCATCAAC AGTCTAGAGG CATATTCTGT	120
20	GTCATTCCCTT CCCGTCTCCT TCATAGAATA CTACTTTTC CTTTTGTCCTC CTGGCCATT	180
	TCCATCATCT GCTGATTATTG GCTAACACCA GGATGCTGGC AAAGCTTACA GTGATAGGCA	240
	CATGTGTTCA GTGATGTCCTA ATACACTCTT ATCACAGTGG TTATTCCTTC TTACTCTTT	300
25	CAAATGCATT ATTCTACCCC TCAACCTAYA TCCAATCATT AGAACTATAC CTGACTGGAG	360
	CCCAGAACTT GGGACCAATA CTTAATTCAA ATAGCAGGGG CTTGCTCACA AACATTAAGC	420
30	CCAAMAAGAA CCACAGCACT TTGAAAAGT CAAATAGGSC TTTGGTAGCT CTGTACATT	480
	NGCAATTTCAC ATTGTTATTA AGTTTATAGC ACTAATAACA CTTCAGTCGT GAATCTACAG	540
	TCTCAATATG ATAAGTCCTA GAACATGTC TAGAAATAGT GGTACCTTGC TGCTATTATA	600
35	CTTAGTAAC TATACCCCAA TATAATAATA AGTATTAAT ACAGATTGTG TATGCATTCT	660
	TTGTGTGTAT ATGCCAACTG TACTACTTAA CCTCACTGAT GAGCAATTAG AAAAATACAC	720
40	AAATGTCACT AGTAAAATA AGTCTGGTC AATTCAAGATG ATACGTGAAC CTGATAAATG	780
	CTCTAATAGA TATGCTATTT TGCTCTGTAT TGCTTGTCTT ACAGTATGGT GCATGTTGTT	840
	TGCTAAGTAA AATGATAATA ATAATAAAAGT ATACCCAATT TTAAGGTAG AATTAAAATT	900
45	TTGCACATAT GCTTCTTGAT ATTCTGAAAT GTATTCTGTG GSTIMATTAT CTTATTCA	960
	CACATTGKMGCTWGCTTTT ACCCCTAGGA AATAACTGTC CAAGTATATA TCTCGTCTTC	1020
50	TTCTGTAA CTTGATTAAC ACTGCTTACT TCAACTTACA ACATGTAAA GCCAGAATAC	1080
	CTCATTTAA CAGTAAAAAA AAATAATTATG ACCTGATGTG TTCTCTGTAA TTTGATTGTA	1140
	ACTACCTAAA TAGGCTTAAC TGATAATAATA AATATACAAT TTTGGAAAAA AAAAAAAA	1200
55	AAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAGGGCGGC	1260
	C	1261

60

(2) INFORMATION FOR SEQ ID NO: 83:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1045 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

TCGAGTTTTT	TTTTTTTTTT	TTTTAAGCAA	CAGTTTATTG	AGACGGAAAA	AATATGATCC	60	
15	AGCAAAGGCG	AGGAGGCGAG	CGGGGCCCG	AGCCAGCTGG	TGTCAATTGTC	ACTGGCTCCC	120
	AAACCTGACT	CCTGTGGACG	TGTCTGTACC	CCAAACACAG	CTGCCACACCC	CAGCCCTGGC	180
20	ACAGAGCCCT	TCTGAAAGAA	AGAAAAAAAGA	AGAAAGACGC	GGCACCTGAC	GCCAGCGGGT	240
	AAAAGCAGGG	CCCCAGAGGC	ATTTATTGAA	AACACAGCAT	CCAAACACAG	ACATCTAGGC	300
	CAGGCCGCGAT	GGTTACAGTG	ATGAGAGGGT	CACTAGACAA	TTATCCACAA	TTCTACGACA	360
25	TGAGACAGAG	ACTCAGCAAC	AGTCACAGAC	AGAAGGGTCA	TGTGTTCTT	CCTGGGCAGG	420
	GCTGAATGTG	GCAGGTGCGG	CGTGGAGGCT	GCGTCCTGGC	GGTTTGCTCC	CAGGCAAGGG	480
30	GTACGGGGGG	CCGGCTTGGC	TGGGTGGGGA	CCTCAAGTCT	GAGGGTGAGG	ATGGCTGAAT	540
	CTACCTCGCT	TATGCTCTAG	GGACGGTCAC	CCATACCTAG	GATGACCCCA	GCCAGACCCCT	600
	AGAAGGTCTG	ATGCCCATCC	CAAGTNCCCC	CCGGAGGAGA	AGAGTTCCCT	GGCAGGGGTG	660
35	ACACATTCCC	GGTCAACAAG	CCACAACACA	GTGGTGCTG	CACTCTCTCA	GCTGTTGCCA	720
	CAACACTTGG	TGCTGGAAATT	TTCTCCACGT	AGTGAAACTT	TTAAGGGACA	CATGAATAAT	780
40	TTAAAAAAAGTC	ACACAAAAGT	CTACGAAAGG	CAGGAATCCT	CACTCTGCTG	AGAGCTACCT	840
	CCTGAGATGT	CGCTTCCCGGA	CCCCGGCAGA	GGGCAGGAGC	GACATCAGCT	CGGCAGGAGG	900
	ATCCTNGCCA	GGCGGAGGGC	TGGCTCTGGT	TATTATAAAT	AATCTAATT	AAATACGCAC	960
45	ATACACACAG	ATGTCCTGCT	TCTACCNAAC	GCCAGAAAAA	GCAGACATTA	GCATCACACT	1020
	GTCAACACTT	CCTCGAGAAC	NGAAG				1045

50

(2) INFORMATION FOR SEQ ID NO: 84:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2877 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

5	GAATTGGCA CGAGACAAGA TGGCAGTCAG CAGCTTCCCA AAAGATAGGG ATTACAGAAG	60
	AGAGGTGATC ACAGACATGA AAAGATGCGA GACGCCGGAG ATCCTTCACC ACCAANTAAA	120
10	ATGTTGGGA GATCTGATAG TCCTGAAAAC AAATACAGTG ACAGCACAGG TCACAGTAAG	180
	GCCAAAAATG TGCATACTCA CAGAGTTAGA GAGAGGGATG GTGGGACCAAG TTACTCTCCA	240
15	CAAGAAAATT CACACAACCA CAGTGCTCTT CATAGTTCAA ATTACACATTG TTCTAATCCA	300
	AGCAATAACC CAAGCAAAAC TTCAGATGCA CCTTATGATT CTGCAGATGA CTGGTCTGAG	360
	CATATTAGCT CTTCTGGAA AAAGTACTAC TACAATTGTC GAACAGAAGT TTCACAATGG	420
	GAAAAACCAA AAGAGTGGCT TGAAAGAGAA CAGAGACAAA AAGAAGCAAA CAAGATGGCA	480
	GTCAACAGCT TCCCAAAAGA TAGGGATTAC AGAAGAGAGG TGATGCAAGC AACAGCCACT	540
20	AGTGGGTTTG CCAGTGGAAAT GGAAGACAAG CATTCCAGTG ATGCCAGTAG TTTGCTCCCA	600
	CAGAATATTT TGTCTCAAAC AAGCAGACAC AATGACAGAG ACTACAGACT GCCAAGAGCA	660
25	GAGACTCACA GTAGTTCTAC GCCAGTACAG CACCCCATCA AACCAGTGGT TCATCCAAT	720
	GCTACCCCAA GCACTGTTCC TTCTAGTCCA TTTACGCTAC AGTCTGATCA CCAGCCAAAG	780
	AAATCATTG ATGCTAAATGG AGCATCTACT TTATCAAAAC TGCCACACCC CACATCTCT	840
30	GTCCCTGCAC AGAAAACAGA AAGAAAAGAA TCTACATCAG GAGACAAACC CGTATCACAT	900
	TCTTGACAA CTCCCTCCAC GTCTTCTGCC TCTGGACTGA ACCCCACATC TGCACCTCCA	960
35	ACATCTGCCT CAGCGGTCCC TGTTTCTCCT GTTCCACAGT CGCCAATACC TCCCTTACTT	1020
	CAGGACCCAA ATCTCTCTTAG ACAATTGCTT CTCGTTTGCA AAGCCACGCT GCAGCTTAAT	1080
	AATTCTAAATG TGGACATATC TAAAATAAT GAAGTTCTTA CAGCAGCTGT GACACAAGCC	1140
40	TCACTGCAGT CTATAATTCA TAAGTTCTT ACTGCTGGAC CATCTGCTTT CAACATAACG	1200
	TCTCTGATTT CTCAAGCTGC TCAGCTCTCT ACACAAGCCC AGCCATCTAA TCAGTCTCCG	1260
45	ATGCTTTAA CATCTGATGC GTCAATCCCCA AGATCATATG TTCTCTCCAAG AATAAGCACA	1320
	CCTCAAACTA ACACAGTCCC TATCAAACCT TTGATCAGTA CTCCCTCTGT TTCACTCACAG	1380
	CCAAAGGTTA GTACTCCAGT AGTTAACCAA GGACCAGTGT CACAGTCAGC CACACAGCAG	1440
50	CCTGTAACTG CTGACAAGCM GCAAGGTCAT GAACCTGTCT CTCCCTCGAAG TCTTCAGCGC	1500
	TCAAGTAGCC AGAGAACGCC ATCACCTGGT CCCAATCATA CTTCTAAATAG TAGTAATGCA	1560
55	TCAAATGCAA CAGTTGTACC ACAGAAATTCT TCTGCCGAT CCACGTGTTCA ATTAACGCCT	1620
	GCACCTAGCAG CACACTTCAG TGAAAATCTC ATAAAACACG TTCAAGGATG GCCTGCAGAT	1680
	CATGCAGAGA AGCAGGCATC AAGATTACCC GAAGAACGCG ATAACATGGG AACTATTAC	1740
60	ATGTCGGAAA TTGACTGATTA ATTAAAAAT TTAAGATCTT TAGTCCGAGT ATGTGAAATT	1800

	CAAGCAACTT TGGGAGAGCA AAGGGATACT ATTTTGAGA CAACAAATTA AGGAACCTGA	1860
5	AAAGCTAAAA AATCAGAATT CCTTCATGGT GTGAAGATGT GAATAATTGC ACATGGTTTT	1920
	GAGAACAGGA ACTGTAAATC TGTTGCCAA TCTTAACATT TTTGAGCTGC ATTTAAGTAG	1980
	ACTTGGGACG GTTAAGCTGG CCAAAGGAA TGACAAGGGG ACGGGGCTCG TGAGAGTC	2040
10	TTCAGGGAA AGATACAAGA TTGATTGTA AAACCCCTGA AATGTAGATT TCTTGTAGAT	2100
	GTATCCCTCA CGTTGTAAT ATGTTTGTA GAGTGAAGCC ATGGGAAGCC ATGTTGAA	2160
15	GAGCTTAGAC ATCCAAAATC AATCAATGCT GAGGTGGCTA AATACCTAGC CTTTACATG	2220
	TAAACCTGTC TGCAAAATTA GCTTTTTAA AAAAAAAA AAAAAAATTG GGGGGTTAA	2280
	TTTATCATTC AGAAAATCTTG CATTTCAAA AATTCAAGTC AAGGCCAGG CGATTGTGT	2340
20	CTAAGGATAAC GATTTTGAAC CATACTGGCA GTGTACAAA TATGAAACAA CTGTTTCCAC	2400
	ACTTGCACCT GATCAAGAGC AGTGCCTCTC CATTGTTTT GCAGAGAAAT GTTTTTCATT	2460
25	TCCCGTGTGT TTCCATTTC TTCTGAAATT CTGATTTTAT CCATTTTTT AAGGCTCCTC	2520
	TTTATCTCCT TTCTTAAGGC ACTGTTGCTA TGGCACTTTT CTATAACCTT TTCATTCTG	2580
	TGTACAGTAG CTAAAATTG CAGTGATTGA GCATAACCTA CTTGTTGTA TAAATTATTG	2640
30	AAATCCATTT GCACCCGTGA AGAATGGACT TAAAAGTACT GCTGGACAGG CATGTGTGCT	2700
	CAAAGTACAT TGATTGCTCA AATATAAGGA AATGGCCAA TGAACGTGGT TGTGGGAGGG	2760
35	GAAAGAGGAA ACAGAGCTAG TCAGATGTGA ATTGTATCTG TTGTAATAAA CATGTTAAAA	2820
	AAAAAAAAA AAAAAAAGGG CGCGGGCTCG CGATCCCTAGA ACTAGCGGAC GCGTGGG	2877

40

(2) INFORMATION FOR SEQ ID NO: 85:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1367 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

	AATCATGAGC CTCCAGAAGA GACAGATGGC CCACCAAGGAG CTGTTGCTCT GGTTGCCTTC	60
	CTGCAGGCCT TGGAGAAGGA GGTCGCCATA ATCGTTGACC AGAGAGCCTG GNAACTTGCA	120
55	CCARAAGATT GTTGAAGATG CTGTTGAGCA AGGTGTTCTG AAGACGCAGA TCCCGATATT	180
	AACTTACCAA GGTGGATCAG TGGAAAGCTGC TCAGGCATTC CTGTGCAAAA ATGGGGACCC	240
60	GCAGACACCT AGATTTGACC ACCTGGTGGC CATAGAGCGT GCGGAAAGAG CTGCTGATGG	300

	CAATTACTAC AATGCAAGGA AGATGAACAT CAAGCACTTG GTTGACCCCA TTGACGATCT	360
	TTTCTTGCT GCGAAGAAGA TTCCCTGGAAT CTCTCATCACT CGAGCTCGTC ATGGAGGCAA	420
5	CGAGCTTGGG ATGGGTAAAG TCAAGGAGGC TGTGAGGAGG CACATACGGC ACGGGGATGT	480
	CATCGCCTGC GACGTGGAGG CTGACTTTGC CGTCATTGCT GGTGTTTCTA ACTGGGGAGG	540
10	CTATGCCCTG GCCTGCGCAC TCTACATCCT GTACTCATGT GCTGTCCACA GTCAGTACCT	600
	GAGGAAACCA GTCGGACCCCT CCAGGGCACC TGGAGATCAG GCCTGGACTC AGGCCCTCCC	660
	GTCGGTCATT AAGGAAGAAA AAATGCTGGG CATCTGGTG CAGCACAAAG TCCGGAGTGG	720
15	CGTCCTGGGC ATCGTGGGCA TGGARGTGG A TGGGCTGCC TTCCACAAACA MCCACGCCGA	780
	GATGATCCAG AAGCTGGTGG ACGTCACAC GGCACAGGTG TAACCGTCCA TGTTCGTGT	840
20	GAGCAGAGTC CCTACCAACG GGCAGGTCTG CATCCGGGA GAATGCAGCT GCTTCTGGCG	900
	ACAATCCTGC TAGTAAACAC TGGTCTTCGG TGAGCAACGA ACACCTCCCT GGCCTGGAA	960
	ACTGCATGCC CACTTTCTGG GAGGGGTTAG TGCAGGTGCC GTGGACAAAG GACAACATTT	1020
25	CTCTGGGCT TTTTAACTTT TATTCTTAAG ACTCTAAAGG CGTTGATTTC AACCCCTCCTT	1080
	CACTCTGGCT TCTTCAGGCA ACCCACGTGG TCTCCTGTGA GAATCTTCTC GACAGTTACT	1140
	TATGGGACA CTGTGAACA ATTAACATGCC AGGCAGAGCA TGAGAACAAA CATTCCAGG	1200
30	CCATGTAGGA TAGGATACTC CAGACTCCAG TCATCCTCCC CCATCCATGG TTTCTGTTAC	1260
	TCATGGTTTC AGTTACTCAT AGCCAACGTGC AGACCGAAAA TACTAAATGA AAAATTTCAG	1320
35	AAATAAACAA CTCTTAAGTT TTAAAAAAA AAAAAAWAA ACTCGTA	1367

40 (2) INFORMATION FOR SEQ ID NO: 86:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1009 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

50	GAATTCGGCA CGAGCTCGTG CGAATTCTC GTGCCGAACT GAAACGTATC AAGAAATACC	60
	TGGGCTTGAA GAATATTCACT CTGAAATATA CCAAGAAACA TCCCAGCTTG AAGAATATTG	120
55	ACCTGAAATA TACCAAGAAA CACCGGGGCC TGAAGACCTC TCTACTGAGA CATATAAAA	180
	TAAGGATGTG CCTAAAGAAT GCTTCCAGA ACCACACCAA GAAACAGGTG GGCCCCAAGG	240
	CCAGGATCCT AAACACACCC AGGAAGATGC TAAAGATGCT TATACTTTTC CTCAAGAAAT	300
60	GAAAGAAAAA CCCAAAGAAG AGCCAGGAAT ACCAGCAATT CTGAATGAGA GTCATCCAGA	360

	AAATGATGTC TATAGTTATG TTTTGTMTA ACAATGCTCA ACCATAAAGT TGTGGTCAA	420
5	TGGAACATAC AGCTTAATAG TTTATGCGTG ATTTTCTCAA AATATTGTA AACTTTGAC	480
	AATGCTCATT AATATTATTT TTTCTATTG TAGACCATAT CTGAAAGAAA TAACATTTT	540
	TAAGGCTCTA CCACATAGAC AATATCATGC TAGAATGTGT GTGTGTGTGT GTGTGTGT	600
10	GTGTGTATGT ATGTATAGGT CGGGGAGAGG ATAGTGGTGG GAACAGACAA ATAAGGAAGC	660
	GGGGAGGACT GGATAATTGG TTTTCCCCCCC TAAGAACATT TATTTACGTC TTAAGAGCAG	720
15	ATAAGTGACT AAGACTGAAC ACATACATTT TGTGGAGTAT ATAGTTTCT TGTAAATGCT	780
	GTTCAATTAT TAATGTAACA GTAGCATCAA AATTTTATTC AGGCTTTAGT TGACTCTTT	840
	GGTCAGTTTT ACAATTCTC CTTAAAAGAT ATTTTGGAGT GATGAATGTA GTTTACTTT	900
20	GTATTGAAAT TTGATTTTC TATTTTTATT TTTTAAATAT TGTATTGIG CACAATGTAC	960
	ATTAATCAT TATTACATGC TTAAAAAAA AAAAAAAA AAAACTCGA	1009

25

(2) INFORMATION FOR SEQ ID NO: 87:

	(i) SEQUENCE CHARACTERISTICS:
30	(A) LENGTH: 1367 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

	AATTCCAAA CAAGGTAAA GGAACCAGAA AAGAAAAAAA ATGTAATAA AGTTATAAAA	60
40	ATAAAGAATT TTTCAGGT TAAAAGCTG AAAAGAAAT AATTTTATAT AAGAAAGAAT	120
	TTTATAATGGT AAATTTAGTC CTAAAATAAA ATAACIGTT GTTTAACAAAG GAGGATGTT	180
	CAGGACAAAC CAGAAAGTCC AAGCATGTCA TGAACATTGG TGTAAGTCAT GATAAGATTT	240
45	TATATATATA TATACACACA CACACACACA CCCCCAAAGC TTTTATATAA TCAAGTGTG	300
	MTATTATTAT TAAGTTTTGG TTTGCTTAGG GAAGAAAGAR CTAATTTITA AAAAATCAAG	360
	GTTATTACAT CCATGATCT TCCTGTGTAT GCTTTAAAG TCCCTGTAAAC ATTGAGITAC	420
50	AGGGCTTAA CTCCGTGTC TGAAAATCA CAAACACTGA TGACAATCAA AGCCTCATCT	480
	TAAGGCCCCG TAGAAGATGC CAATCAAAAT AAACIGCATT CCTGAGGCAC TAGGCAAGAA	540
55	ATTAAAGCTA TTCAACTCCT CAAGGCCAG GGACTATTGC GGAAGAGGTG GGCGCGTAAG	600
	ATTGTAAGGG CGGATTTGA AAGATCCAGT AAGTTCAAGT TCTCTATGAA CTAATCATTC	660
60	AAGTCAAAGG CACACTGATG CAAAATCAAGT ATATGGACCC CTGTGTCTGA TTAGCAAGGT	720

	TTCTTGAAG CATTAAACAA CTCCTTCATA AAGGTTATAA AAGGCTTATG GRAGTTATAT	780
	TTTATAATCA AGATTAAATC TTATAGTTG TTTACAAAAT TTTGAAAATC AAATGTGATT	840
5	GGCTTCAGGC TGTTTTTATT AGGGCTTCTT GTTTAGAAAG TTAAGTCACC TCTCTCAAAG	900
	AATGAAGGTT TTGCTTTTT TIGAAATCCT TGAAATTATCA CTTGGRTTAA ATAAATGACT	960
10	TTACGATGAC CTGTAATTTC ATTTTGTAA GTCAAGTGT TTAAACCTTT TGTATTTGAC	1020
	AAGCTTICCA AAATCAAATT ATAAATTATG TATTTTCTA ACCTAATTAA TCCTTTAAGA	1080
	TCTTAGTTTC CCTAAAGTCC TAAAATGACA TAATTTGGCT TATTTGGTAT AAAAATTATA	1140
15	TAGGAAACAT TGTCAAATGT GAAATGGTGT TTGGTTTCT TTGGGCTGTA TTGTATAAA	1200
	TATGTTATTG GTGTATGTC CAAAATTATG TGAAACTCCT ATAATTCTAA TATAACTTAG	1260
	TGTACATTAT CAGTAATAAT CATAATTGTT ATATTTAAAT TATTGTGTGC CACAGAGGTA	1320
20	AAAAAAAGG AATTGATAT CAAGCTTATC GATACCGTCG ACCTCGA	1367

25

(2) INFORMATION FOR SEQ ID NO: 88:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1088 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

35	GAATTCCGCA CGAGTGAAT TTTGTCGATT TCAAAATGG AAAATACATA ATATGCCAGG	60
	CACTTCCTGG GCAATACAGA TACCTGCAGT AATGGAGTGA GCACCAGCAT CTTCCCTGAT	120
40	GGCGTGTGCA GTGAGGTGAC TCGTCTGTAG TGTCTCAAG GTCACGTAGA GAGCATAACAG	180
	TAAATACTTG TTGACTCTTT CAAACITAAG TTAATGATAC AGTCAGGACT GATAGCCATT	240
	TTGTTGTCCTT TCTTGAAAGT TTACGTGGAA GGCAGACCTT GTGTATGCTT TTCAAAGGGG	300
45	CTCMTTAGC GCACTTGGCG CTTAAGAATT TGAGATCAGT AAGTGTGATG GTCCTAATCT	360
	TTTTTTAAAA GTATTGGAAG TTTGAACYCM CCTGATGGGG TTGGTTTTTT TTTTTTTTT	420
50	TTCCAAAAAA ATAATCATTC AAAATAATCG GTTAACATT TCAATAAGAG CATTACATAC	480
	AAGGAGTTAG GGAACAAAGA GTTTTAAAT CTGGCTCTTT TTATCTCTAC TTAGGGCGTG	540
	CATCTCTCT TCTTACCCCA ACATATACTG ACTTTTCTAG ACCTCCTTTA GGGAGATCTC	600
55	AATATCCCGA ATTTTCTGT GTGGAGAGGG GAAGGAATAT GTCTTTTTT GCTTTGGTCA	660
	GAGTGGATAC ATTTTATAGT TTGTTTTTC AAAGACGGGT CTTCTGAGTC ASTTCTTTCA	720
60	CTGCTGCCGT AAAGAAACTG TATAAAGGTG ATTGAGCACT GAAGGCATGG ATAAAAGGGG	780

	AAATATTCTAG CAGTTCTGAA CGTGCATGTC ATCAAATATA AAGGAGTGAG AACTTGATGT	840
5	ATAAGAAAAA ATGGAAGTTA AAAAAAAWAA AAATCCAAGA ATGGGCTGCT TGTTGCAGTA	900
	GTGAACTCCT CGCTGGAGGT ACTAGAGCGG AGTCTGTCTC AAGGATGCTA TTGGAAGCAC	960
	CCCAGCTGTG GGTGGAAAAC TGCACTTTCT GAGCCTAGTC TTTTATAGCC TGGRGTTTT	1020
10	GATGCTGATG CTTTTACTAC TTGTTCTTAG ACTWTTTTGC CATAACGCTGC TCTGTTTCT	1080
	CACCTCCA	1088

15

(2) INFORMATION FOR SEQ ID NO: 89:

(i) SEQUENCE CHARACTERISTICS:

20	(A) LENGTH: 1861 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

	TCTCTGCCCT TCATCTTGGT AATTAGCCAG CCTCAGATAC TTCTGTGGC CCTGAAGTGG	60
30	ACTCTCAAGG TCAGACCAAG GTTGTGATC TCAGTCCCAC TGTCTTCAGC CAGCTGAAGC	120
	TGTGGGGCTG GGCTGGCAGC TTTATTGTCA TCTTGCTTCA CCATTTTTT TTCTCTCT	180
	TTTCATTCTA TTTTAAGTTT AGACAAAAAA AATACAGAGT CATCCCCAAC CCCCACCCCT	240
35	CTAGAGACCC TCCAGCTAAA AACAGAGCCT GAGTTCAAGG ACCCAAGTGG TGAGCGGCGT	300
	CTTTTGGGG TGAGGGAGCT TGGGTAGATG AGGCTCTGG CTGAGCCCTC CCTGTGGTGA	360
40	TCCCAGCTA AGATGGCCCC TCCTCCCTCC TGGTGGGAGA CAGAGGACTG GACCCCTGGT	420
	CTCAGGTTC AGCAAGTCAG GCTAGGGACC TGGGGGGAGG AGACCCATGG ACTTCACCCA	480
	TACTCAGTGA GGGGGCTCCT GCGGTCTGA CGCCACCCCG CCCCATCAGC ACTTAAGCCA	540
45	CATGACACAA AGTCTGTACC CCACGGAAA TGTTCACCGG CCTGGGGCGT GTGCATGGCC	600
	TCCCGGGCTG TGGGGCAGCC GCATCTGTGA GGTGACYCGT GAAAGTAGGT GATTCCYTTG	660
50	CAGAACTTCA GGGACTGGGA GCAGAGGCC CTCACTAAC GACGTTGTG CGACATAGTA	720
	TTGTATCCAC CTTAGTATTG TATCGAGCCT TTTCTGTGTT TTAATGAGAA AGCAGAACAC	780
	TAGTTTCTA TTTAAGACTT TAAGGTTTG TGGGGGGGG CGGGATTAAC ACAACATTTG	840
55	GCTTGTGTTT CTTTTCTT TGATTTCCAC ATCAGGTGTG TGCGAGTGTG TGTGTGTGGA	900
	GATGTTAAGA GCCTCACAAAG GAAACTGGGT TATTGGAGGC CAAGGCGGCT TACAGTTCTC	960
60	TGGCTTCGTC ACTTAATTCC TGAATGTTTC AGAGAAACAG GAATCAGAAA ATAGCAGATA	1020

	TCATGTAGGA AAGAGAGGAT AAACAAAGAA AAAAGAAAAA AAAATAAGCT CATAACCAA	1080
	TTCACAAAGC CTATTTTTA AACCAAAGCA CATTGAT GAGTATGGAA CCTCCATGG	1140
5	CTCAGAAAAA AGATGCTAAT ATATTTATCT CATTGTTAC ATAAGCTTT ACAGTTTCAG	1200
	ACCTCAGCAG CTGTAAGGCC AGTCCAGGG ACCCTCCCT GCTGCTGGAA ACCCTCTGA	1260
10	GTTGGCCCTG GAGTGGCTCA SGGGCAGAGA AGGGTAGCCC TGGGGCTGGG GGAGGGATTG	1320
	GAAGCCTCCC TGGAGTCACC TGAGCCCTCG TCCCCATTCC CAGGGCCCT CCAAGCCCAG	1380
	CTGGCACCAA ARAGCTTGGG CCCGTCTGA CCAGCCCCA AGGCCCTCTG GCGGACCAT	1440
15	GCTGGTCTG ACCAGCTAGC CTACGGGGG ATGGCCGTCA GTTCTGGCA CAGGACCCGA	1500
	GTCCTGGCTT GGGTCCCCCT GCTGCTCTGC CGGTGACCCCT TGGGGATGGG TTGATGGAG	1560
20	GGTCCCACTC AAGCCAAAAA GCGGGGACCT TTGCGCAGCT CTGTCGACTC TGGTGGGTCC	1620
	CCACTCCTGG GGCCCCCTAA CCCACCCCCA GGCAGCGAA GGGCTGACT GGGCTGGTC	1680
	CTTACCAACA TAGACGGTGC AAACACTCTT AACAGTGTG TTTTTGTATC AATATGTTG	1740
25	TGCAGTGTG AAATGTTTTA TTTCTCAGAC TTGGGGCGAG TGACCGGGTG GCAGGCCGGC	1800
	TCCGCCACTG CAATGCTCCC GCGGACCGA GCCCCAGCAA GGGCTCCTCC AGGATTGCAA	1860
30	A	1861

	(2) INFORMATION FOR SEQ ID NO: 90:	
35	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1259 base pairs	
	(B) TYPE: nucleic acid	
40	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:	
45	AATTGGCAC GAGCTCGTGG AGAGATTGAA GATGGCGGT TCTCAGGGGG TGGAGGAAAT	60
	GCGGACCGCG TGTTCTGGG GGAGTTGGG GTCGCAATG TCCATACTAC TGACTTTCCC	120
	GGTAACTATT CCGGTTATGA TGATGCCTGG GACCAGGACC GCTTCGAGAA GAATTTCGGT	180
50	GTGGATGTAG TACACATGGA TGAAAATCA CTGGAGTTTG ACATGGTGGG AATTGACCCA	240
	GCCATTGCCA ATGCTTTTCG ACGAATTCTG CTAGCTGAGG TGCCAACATAT GGCTGTGGAG	300
55	AAGGTCTGG TGTACAATAA TACATCCATT GTTCAGGATG AGATTCTTGC TCACCGTCTG	360
	GGGCTCATTC CCATTCATGC TGATCCCCGT CTTTTGAGT ATCGGAACCA AGGAGATGAA	420
	GAAGGCACAG AGATAGATAAC TCTACAGTTT CGTCTCCAGG TCAGATGCAC TCGGAACCCC	480
60	CATGCTGCTA AAGATTCTC TGACCCCAAC GAACTGTACG TGAACCACAA AGGCTGATCT	540

	MTTTCAGAG GGCACATATCC GACCAGTCCA TGATGATATC CTCATCGCTC AGCTGCGGCC	600
5	TGGCCAAGAA ATTGACCTGC TCATGCACTG TGTCAAGGGC ATTGGCAAAG ATCATGCCAA	660
	GTTTCACCA GTGGCAACAG CCAGTTACAG GYTCCCTGCCA GACATCACCC TGCTTGAGCC	720
	CGTGGAAGGG GAGGCAGCTG AGGAGTTGAG CAGGTGTTTC TCAMCTGGTG TTATTGAGGT	780
10	GCAGGAAGTC CAAGGTAAAA AGGTGGCCAG AGTTGCCAAC CCCCGGCTGG ATACCTTCAG	840
	CAGAGAAATC TTCCGGAATG AGAAGCTAA GAACGTTGAG AGGCTTGCCC GGGTTCGAGA	900
	TCATTATATC TTCTCTGTTG AGTCAACGGG GGTGTTGCCA CCAGATGTGC TGGTGAGTGA	960
15	AGCCATCAA GTACTGATGG GGAAGTGGCG GCGCTTCTTG GATGAACTAG ATGCGGTTCA	1020
	GATGGACTGA GCTTGGATGC TTCTGAGGCA AGCTGAAGCT TTGGGTTCTG ACTGACCCAC	1080
20	CCTACAGGAC TGCTGAACAG AGAGCCCCAGT GTGACTAGGG ATCCTGAGTT TTCTGGGACA	1140
	ATTCCAGCTT TAATCAATAC ATTTTGTTAA ATGTGCCATA AAATGAGACT TTTTACGCCT	1200
25	TTATAAAGGCC TTAGATGTAA ATAAAATCAC CCAAAACAAAA AAAAAAAA AAAACTCGA	1259

30 (2) INFORMATION FOR SEQ ID NO: 91:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1566 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

40	CTAGAAGAGC AAGCCCGCCA GNANTGATGA AAAC TGATTT TCCTGGAGAC CTTGGCAGTC	60
	AGCGACAAAGC TATTCCAACA ACTAACAGAGAT CAGGACTCCA GTAGCAGTGA GTTCTGCACC	120
	TTCTGGTGAC AGTGAGGGTG ATGAAGAGGA GACGACACAA GATGAAGTCT CTTCCCACAC	180
45	ATCAGAGGAA GATGGAGGGG TGGTCAAAGT GGAGAAAGAG TTAGAAAATA CAGAACAGCC	240
	TGTTGGTGGG AACGAAGKGT TAGAGCACGA GGTCACAGGG AATTGAAATT CTGACCCCTT	300
	GCTTGAACTC TGCCAGTGTC CCCTCTGCCA GCTAGACTGC GGGACCGGGG GCAGTTGATT	360
50	GCTCACGTGT ACCAGCACAC TGCAGCA GTGAGCGCCA AGAGCTACAT GTGCTCTGTC	420
	TGTGGCCGGG CCCCTAGCTC CCCGGGGTCA TTGGGTCGCC ACCTCTTAAT CCACTCGGAG	480
55	GACCAGCGAT CTAAC TGTC TGTGTTGGA GCCCCGGTCA CCAGCCATGC CACTTTAAC	540
	AGTGAGAAC TTCTGAAGT ACTAAATATG GAATCCCTAC CCACAGTCCA CAATGAGGGT	600
60	CCCTCCAGTG CTGAGGGGAA GGATATTGCC TTTAGTCCCTC CAGTGTACCC TGCTGGAATT	660

	CTGCTTGTGT GCAACAACTG TGCTGCCTAC CGTAAAMTGC TGGAAAGCCA GACTCCCAGT	720
	GTASGCAAGT GGCGCTCTACG TCGACAGAAT GAGCCCTTGG AAGTACCGCT GCAGCGGCTG	780
5	GAACGAGAGC GCACGGCCAA GAAGAGCCGG CGGGACAATG AGACCCCCGA GGAGCGGGAG	840
	GTGAGGGCGCA TGAGGGACCG TGAAGCCAAG CGCTTGCAGC GCATGCAGGA GACAGACGAG	900
10	CAGCGGGCAC GCCGGCTGCA CGGGGATCGG GAGGCCATGA GGCTGAAGCG GGCAATGAA	960
	ACCCCGAAA ACCGGCAGGC CGGGCTCATC CGAGAGOGAG AGGCCAAGCG GCTCAAGAGG	1020
	AGGCTGGAGA AAATGGACAT GATGTTGCGA GCTCAGTTTG GCCAGGACCC TTCTGCCATG	1080
15	GCAGCCCTAG CAGCTGAAAT GAACTCTTC CAGCTGCCTG TAAGTGGGGT GGAGTTGGAC	1140
	ARCCAGCTTC TGGCAAGAT GGCTTTGAA GAGCAGAAC GCAGYTYTCT GCACTGAACC	1200
20	ACACCCCTCCT GCCTGCCCTC CTTCCCACCT ACCTACCCAC CCACCCACAC CCACAGCCAC	1260
	GAGGACCAAGT GCTGCTGCCA CCCACGAGGC CCTGTCCCTG CTGCCAGAGG CAGGCCTGGG	1320
	TTTATTGCAG GTGGACCTGA GCAGCCCTTG CATATGGAA CAGGATGATG GGGTCAGGAG	1380
25	GGACCTGGCT CAAGGCAGCT CTGGACAAGG GAGCAGGCAG TCCAGAGAAC TGGCTCCCC	1440
	AGCCCACTGC CACAGGCTGT GCTTCTAGGA CTGTGGGCCCT CTGTGTGGCC CATGAAGTTG	1500
30	TGAAGTCAAA TAAATTAATT TTATCTTAA AAAAAAAA AAAAAAYYGG GGGGTTTTTT	1560
	TGGGGG	1566

35

(2) INFORMATION FOR SEQ ID NO: 92:

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1593 base pairs	
40	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:	
45	GGCACGAGCC TCGGCCTCGG TGGCGGTGGT GGACACGTCG AGCCGGTAG AAGTGGAGGG	60
	GCCGTTGAA GAGTCGTGAG GGGGTGACGG GTTAAGATTC GGAGAGAGAG GTGCTAGTGG	120
50	CTGGACTTGA CCTGGAAAGA ATCTCTGCT GACTCTAAC TTTTCTGGA AAAATGGAT	180
	CATCCCCACC ATATGGGGAT GAGCTATATG GACTCCAACA GTACCATGCA ACCTTCTCAC	240
55	CATCACCCAA CCACCTCAGC CTCACACTCC CATGGTGGAG GAGACAGCAG CATGATGATG	300
	ATGCCTATGA CCTTCTACTT TGGCTTTAAG AATGTGGAAC TACTGTTTTC CGGTTGGTG	360
	ATCAATACAG CTGGAGAAAT GGCTGGAGCT TTTGTGGCAG TGTTTTACT AGCAATGTC	420
60	TATGAAGGAC TCAAGATAGC CCGAGAGAGC CTGCTGCGTA AGTCACAAGT CAGCATTCCG	480

5	TACAAATTCCA TGCCTGTCCC AGGACCAAAT GGAACCATCC TTATGGAGAC ACACAAAACT	540
	GTTGGGCAAC AGATGCTGAG CTTTCCTCAC CTCCCTGAAA CAGTGCAGCA CATCATCCAG	600
	GTGGTCATAA GCTACTTCCT CATGCTCATC TTCATGACCT ACAACGGTA CCTCTGCATT	660
	GCARKAGCAG CAGGGGCCGG TACAGGATAC TTCCCTTTCA GCTGGAAGAA GGCAGTGGTA	720
10	GTGGATATCA CAGAGCATTG CCATTGACAT CAAACTCTAT GGCGTGGCCT TATCGATTGC	780
	AGTGGGAAGT TGTGAAAGAC TTGAAAGACGT GATTCCCTGCT CCAATCATCC CTTCTTGCTC	840
	CTCTTTGKGC ACGTACACAC ACACACACAC ACACACACAC ACACACCCGT GYTCAAACAG	900
15	AGGTTTAGTT TACAGTCTCT GAACCTAAAGT AGTAACCTCC CAAATTGTTT TTCTTAATAA	960
	GCTGAGATTG CCATTTCTCT TAAGGAGAAG CCACCCATGA GATGCTTTT CCTTCTCCAT	1020
20	CATCTTAGAG CCAAGTTATA TGTTCTTGTC TAATCCATGT AGCTTTTGT TCAATGACTT	1080
	GATCATCTGC TTCCCTTTTG AATTTTTAAC AGATAGTAAG TAAATTGTT GGTTTTTCC	1140
	CCTGGGTCAG TGATGGAAAG GGGTTAACTT CAGCCAGGAT TGATGGCAGC TGAGGGAAAT	1200
25	TCTTGCCCAA CTAAACCCAG AACTCAAAC TAAACATTAGA AAATAAGGTC CAGGGCCGA	1260
	CACAGTGGCC CAAGCAAGTA ATCCCAGCAC TTTGGGGGC CAAGGCAGGC TGGATCACCT	1320
30	GAGGACAGGA GTTCGAGACC AGTCTGGCCA ACATGGGAA ACCCCGTCTC TACTAAAAAT	1380
	ACATAAATTAA CCCGGGCATG GTGGTGGGGC CCTGTAAATCC CAGCTACTCA GAAGGCTGAG	1440
35	GCAGGAGAAT CACTTGAACA TAGGAGGCGG AGGTTGCAGT GAGCCAAGAT GGCGCCATTG	1500
	CACTCCAGCC TGGGTGACAA GNGTGAACACT CCATCTCATA AAAAAAAA AAAATANTCG	1560
	AGGGGGGGCC CGGACCCAAA ACGCCGGAAA GTG	1593
40		

(2) INFORMATION FOR SEQ ID NO: 93:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 970 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

	CTCGTGGCGA ATTGGGCACG AGGTGGCCAG GCTCTCAGGG CAGAGGGTCC AGTGTGATCA	60
55	CTTGGCATGG CCTCTCTCCC CTCCCTGAGCT TGTGCCAGGG CCCCAGGGCT GACCTGGAGA	120
	GGAAAAWGGC AGAGGGTGAA GATGGGGTGT CTGGTTTGGG GACCATCCTG GCCCCCCTTG	180
60	TCACTGTTGG CATCTCTCT GCACAGTGGC ATTGCTGGGA GGTGCTTACT GTGCCTATTG	240

250

	AAGGGGCTGG CAGCCGCAGC CTCACTGCAG ATCAGGGACT TGGCTTCCCG GTTGACCACA	300
	GGTCCAAGAA CCTGCAGGGT CCAGCCTCCC CCCCCATCCCC AGTCTTCCCC ACCCTGGCCC	360
5	GGCCCTCCAG GTGCAGAAC ATGCAGGCC CTCCTCCAGGA CTGTGGGAGG AGTGTGTCCC	420
	TCAGACTGCC CTGTGTCTTG GCTCCCTTTA CCACCTCTTC CAGAGGTTGT CACCTGCAGC	480
10	TGCCCCAGGA TAAAGGCAAG GCCAGAGAGG ACTCCTGAAC TCCGTGTGTC CTGGGTGGC	540
	AGGGCAAAC ATAGCCAAT GGTGGCTGA GGGGGCAT GGTGARGACA CCCTTGGTGG	600
	CTTGTCCCAC ATCAAGCTGG GARGTACAC TGAGGATGCA TTAGTCTGCA GCGTATGATA	660
15	AAAACGGCAT TTCAGGCCAG GCGTGGTGGC TCATGCCGT CACCCAGCA CCTTGGGAGG	720
	CCGAGGTGGG CAGATCACAT GAGGTAGGA CTTTGAGACC AGCCTGGCCA ACATGGTGAA	780
	AACTCATCTG TACTAAAAAA ACAAAAATTA TGTGGGTTGG TGGTGTGTGTC CTGTAATCCC	840
20	AGCTACTTGG GAGGCTGAGG CAGGAGAAC ACTTGAACCT GGGAGGGGA GGCTACAAACG	900
	AGCCGAGATT GCACCACTGC ACTCCAGCCT GATCCGTCTC AAAAAAAA AAAAAAAA	960
25	AAAAACTCGA	970

30 (2) INFORMATION FOR SEQ ID NO: 94:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 934 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

40	TCTCTCTCTC TCTCTCTCTC TCTGCTGTAA AGAACTCCCA AAACTCAAAT GTATCAGGAA	60
	ATGTAAGGT TAAGTCTGAC TACAAGAAGG CAAAAATTGC ACCAGCTTCC TAAGTGAAGA	120
	ATAATAGAAT AAAACATATA GAGGGCAGAA ATAAAATGAG GTGTATCTGG AGAATTTCAT	180
45	GATGAGCATT TAGATTTAGC AATGCCAAT GTCATGCTGA CACTGTTGT CATGACCTTG	240
	TCTTCAGCTA GTAATTTGGG GTTGTACTTT TTTAAATTAA ATTTGAATG TTCTTCATG	300
50	TTTGGTACCT CTCTCCTCAC TGCTAAAGAT AAATTGTTA TCTGTATAAC ATAACACAC	360
	CAATGTCAATT TATGTATAC GCTAGTACAC AAATGTGTTT TTTTATAAG TAATGAARTA	420
55	TTTGTGTGA AAAATGTATT ATTTGTGCCA CCGTTTATAT CTGTGTTCAT TTTCTGTGTG	480
	TATATGCGTG TGTATTCGAA TCTCAATTTC TCTTTTACTC TAGTTAGAT TAAGACATAT	540
	TTAGATGAAA TTTTAAAAAT AACATTGGAA ATAGGAGGCT AAGTTTGTT SAGTCTCATT	600
60	CCCTTGGGG GAAATTGCTT TTGCCATTTC ATTTCATGT ACAATAACCT AAAAAGGATC	660

	TCCTACTGAC TTCCCTTCCTA ATTATTATTG TTTTACACGA AAGAAAGGAA ATACGTTTC	720
5	AATTGAGTTG TTGAAATCA TTCACTTTGT GTAGATTTCC CAGACTGATG TTTCATTTGTA	780
	AGAATATTAC ATTATAGACA GGTTGCCAT TTCACAAGCA ACTAATCCAT AGTTTGGAA	840
	GCCCCTTTA AGAGACCTGA ATATCTTTGT TTTAATAAA ATACTTAGAG TTTAAAAAAA	900
10	AAAAAAAAA AAAAAAAA AAAAAAAGG TAAA	934

15 (2) INFORMATION FOR SEQ ID NO: 95:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1392 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

25	CAGCTCAGCT CTGCGCTGCT GCACGCCAAC CACACACTCA GCACCATTGA CCACCTGGTG	60
	TTGGAGACGG TGAGAGGGCT GGGCGAGGCG GTGAGGACAG AGCTGACCAC CCTGGAGGAG	120
30	GTGCTCGANC CGGCCACCGGA GCTGGTGGNT GCGGCCGAG GGGCTCGACG GCAGGCCGAG	180
	GCTGCGGGCC AGCAGCTGCA GGGGCTGGCC TTCTGGCAGG GAGTGCSCCT GAGCCCCCTG	240
	CAGGTGGCTG AAAATGTGTC CTTTGTGGAG GAGTACAGGT GGCTGGCCTA YGTCCCTCTG	300
35	CTGCTCCCTGG AGCTGCTGGT CTGCTCTTC ACCCTCCTNG GCCTGGCGAA CAGAGCAAGT	360
	GGCTGGTGAT CGTGATGACA GTCATGAGTC TCCCTGGTCT CGTCTGAGC TGGGGCTCCA	420
40	TGGGCTGGAA GGCAAGCCACG GCGGTGGGCC TCAGTGACTT CTGCTCCAAT CCAGACCCCTT	480
	ATGTTCTGAA CCTGACCCAG GAGGAGACAG GGCTCAGCTC AGACATCTG AGCTATTATTC	540
	TCCTCTGCAA CGGGGCCGTC TCCAACCCCT TCCAACAGAG GCTGACTCTG TCCCAGCGAG	600
45	CTCTGGCCAA CATCCACTCC CAGCTGCTGG GCCTGGAGCG AGAAGCTGTG CCTCAGTTCC	660
	CTTCAGCGCA GAAGCCTCTG CTGTCCTTGG AGGAGACTCT GAATGTGACA GAAGGAAATT	720
50	TCCACCACTT GGTGGCACTG CTACACTGCC GCAGCTGCA CAAGGACTAT GGTGCAGGCC	780
	TGCGGGGCCT GTGCGAARAC GSCCTGGAAG GCCTGCTCTT CCTGCTGCTC TTCTCCCTGC	840
	TGTCTGCAGG AGCGCTGGCC ASTGCCCTMT GCAKCCCTGCC CCGAGCSTGG GCCCTCTTCC	900
55	CACCCAGGAA TCCAAGCGCT TTGTGCAGTG GCAGTCGTCT ATCTGAGGCC CTCCTCCCGG	960
	CTGGACTGGA GCCTGGCTCC CCTCTTCGTT CCTTCCCTGG CTGCCGGAGA GACCCCACTA	1020
60	ACCCAGCCTG CCTGGCTCT GACCACTAAC ACTCTGGCC ATGGACAGCC TGCACAGGAC	1080

	CGCCCTCCCTG CTCTTGGCCA CTGTGCTCCC ATTTCTGTCC TTGGCCTTGG GAGTAGCTGA	1140
	GGGGCCAGAC TAGGGAGTAG CGCTGGCAGG GGAGGGGGCA GACAGCCTCG CCTCGCACCC	1200
5	TTCATCCCTG GCTGCCGGTC CCATCCTTGG AGGGACTAAAG CTGGGGGTGG GACATGAGTC	1260
	CCCTGCTGC CCCGCCACA TCCCAGTGGG CTCTGACCCC CTGATCTCAA CTGTTGGCAC	1320
10	TAACTTGGAA AAGGGTTGAT TAAAAATAAA AGGGAAGACT ATTTTACAAA AAAAAAAA	1380
	AAAAAAACTC GA	1392

15

(2) INFORMATION FOR SEQ ID NO: 96:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1963 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

GGTANCTGCA GTACGGTCCG ATTCCCCGGT CGACCCACGC GTCCGGAGAA ATGCAAATTA	60
AAACAGTAAA GTGTCACTTT CACTTCTGG ATTGGCAAAG GGTTTTATGT ATTTTACTGA	120
30 CAGTGCTCAA CATTAGCAGT AAACAACAAA TGGTGACTAA ATATGAGCTT CGGAACCTCA	180
GGGAAATGAT CTCCTTATTT CAACCTGCAG ATTCCCTTCCT ACAACCAGTG TAGAGCAGAG	240
TACCAAGGACG GGCCATTGAG CACCCCTGGTG TTGAGATCAA GTGGCCTCTA GTCAGAGTTG	300
35 GGTCAAGGCC ACTGTGAGTG GGCTGCCCTT AACATGAGTC AGCTGTCTAG GACTAGTTA	360
TCTCTGCTTC TCACTTTACT GGTATTATGG GGCAGCTCCT GCTGTCTTCC AATTTGGTGT	420
40 CTTCCAAATC GGCACCGTCT TTTAAAGTTG AGTTTCTTGT TATTCTCACC TGATATAACCT	480
TATTTATCCC ACACCCACCC CAATAACATA TCGTGCTCAG TGTTATCTTT GAGACAACAC	540
45 TTGAATTITA CTCAGCCTGG AGCGCTCTTC ACATGTCTTG TCCAGATCCA GTTCGGACTC	600
ATTCTTCAGC CGTGCATCAG TAAATGGGG CTAGGTTAAA CTGTGGTGAC AAACAACCTC	660
CAAATTTCAG TGGCTAAAA ATCTCTTCC TCATTTATWT ACATTTCACT ATGGGTCAAGG	720
50 TGAGAGGTAG CTCTGTGCTG TGTCACTCTA ACACAGGAAT CCAGACGGAA GGAGGGACAA	780
TCAATAAGAT CCCCATTGCT ATAGAAAAGA RAAAAAAAGTA TGCGGAATAR CACTCYGTTT	840
55 CYTGGAGAWT YCTCTGAAA AAGTCACATG TTATTTCTTC TCACCTCCAT TGGAAAAAAA	900
AAAGTCATGT GGCCATGTGA AAATGTAAGT AGGCAGGGATG GAACAGTCAG AATGCATTCA	960
TAAAATAATGA ACTGAAAATA TCTGGAGAAC AKCACCTATG ACTACCACGA ATGCCAACAT	1020
60 GCATCCCTAA CAACCCAGTG CTGTCACCCCT CCAAACTTTT TATGTCCTTGC AAAGTATTAG	1080

	AACTCTTAT CTGAAGCCAT ACCACTCAGA GGGAAANGCAA AATACATATT GACATCTCCT	1140
5	TTAGGATGTC CTTAGAGAAT TCAAGGAAAAA GAAGTTAAAT AATTTTAAAG TGCTTTGGG	1200
	TACAGCTATT TAGCACTAGA GGGTAAGATT AGACATAGAT TGTAAAGATA ATNATAGGGT	1260
	TAGGGATAGG ATTAGGATCT GGGTCAGAGT CAGGSCCAGA AGTATGGTTA GAGGTGGGT	1320
10	CATGGTCAGG GTSGAGATCA AAGTCAGGGT CAAAGTAAGG GTCAGAATTA GGGACCCAGG	1380
	ATAGGGATCA GGATTTAGGT TCAGTGTCAA AGTCTTGGGA CAAGCTTACGG GTTACAATTA	1440
15	GAACCAGAGC TTTGTTCTCC TCAGGACCCA CCCGAGGGTG GGTCAACCATG GCTTTGGAGC	1500
	GCCTCGTAGT GTGGTGTGTC CACAGKGAAG ACCAGAGTTT CATTGTCCTT AAGACTGACY	1560
	TGGGGAGATG TGGCTGTAGS CCATTGAGGA AGGTGAGGCA ACAGCTTCTC GTCTGCTYCC	1620
20	CCGTGTGCTG AGGAGGGAGT TCTGCCATGG GCTTTACITTT CACATGTTAT ATTCCACAAG	1680
	TCTTGTTTTA CAAAAGCATC CCTTCCTTGA GGCTTCGGCT GCTCATCGCT GCTCATCATM	1740
25	ATAGCGTGCC ATAACATATA GTAAGATTTG GGTGGTGTTC TGGGGAGATA TCTTGGTATA	1800
	GAGAAAGGAG AAATGCTTAG AGCCACCATC AGGACAGTTG GGATGAAAGT TGGGTATAGG	1860
	CAGAGGCTGG AGGAAACATG TGCACTCCCT GTAAACACTT TTATTCATGT TTTAATTACT	1920
30	CATTTTCTT ACAGTGTAA ATTAGTAAAG ATAGTATTGA AAA	1963

35 (2) INFORMATION FOR SEQ ID NO: 97:

	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 1052 base pairs	
	(B) TYPE: nucleic acid	
40	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:	
45	TCATTAACCTT CAGACAACAT CATAAAGCAA TGATAGCTCT TTTCTTGTG ACCACAAYCT	60
	TAACTTGAGC TTTGCTGGGT GTTTTGACACA TAACAATGAG GGACTATTAG ACATAACATA	120
50	ATTTTCATAG GTCATTGCCG TGTCAATGAT AGAGAAGATA ATTGCMAGAK AGTTWATTTC	180
	TGGTGTGTGT ATATGTGCAC AAATGTGCAG GGCCTCTACT TTGCAACTGG AATTTATAGA	240
	CTAATGATAA AATATATCCC TTTAAATATA CAAATGACAA TTGACTTCAA ACTTTCCCAA	300
55	GCCCCACATAG AAATTCCCTG AAAACATATA AAATATTGAG TTCTTCAACC TCAGCACTAT	360
	TGACATTTTG GACCARATAG TTCTGTWIGT KAAAGGCKGT CTTTGCACTG TAGAATGTTT	420
60	AGCAATATTG CAGGCCTCTA TCCACCTGAT ACCGGGCCTG TATCCCCCTG ATACTGGTAG	480

	TTCTTTTTTC CCCCATCACA AATTGTGACA ACCCAGAAAT ATCTCCTTAT ACCTTCCAG	540
	AATGTTTTCC CTGGGGGACA AAAAGCACTC CCATTGAAAA ATCCACTGGT CCCAAATGGT	600
5	TAAAAATTGG TTCCCTTCCC ATTCCTTTA CCAGGTTTGG GGCCAAGCCC CCTTCCCTTA	660
	ATTCCTCTCC CGAAATGAAC TGAAACCCAA CTGTWACTCT TAATGAAATA TTGAAGGKTT	720
10	GAAGCTTAA AAAAAAAA AAAAKTACAG CTTGGCTGG TGCAGTGGCT CAAGCCTGTA	780
	ATCCTAGCAC TTTCGGAGGC CAAGGTGGC AGATTGCCCTG AGCTCAGGAG TTCGACACCA	840
	GGGTGGCAA CATGGTAAA CTCTGTCTCT ACTAAAATAC AAAAAGTTAA CCTGGCATGG	900
15	TGGCAGGTGC CTGTAGTCCC AGCTACTAGG GAGGCTGAGG CAGGAGAATT GCTTGAACCC	960
	AGGAGGCAGA GGTTCAGTG AGCCAAGATT GCCACTGCAC TCCAGCCTGG GCAACATAGC	1020
20	AAGACTCTGT CAAAAAAA AAAAAACTC GA	1052

25 (2) INFORMATION FOR SEQ ID NO: 98:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 929 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

35	ATCCATCACA GCCTTCTAT CTAGGCCACA CTATAAAATC TGGAGACCTT GAATATGTGG	60
	GTATGGAAGG AGGAATTGTC TTAAGTGTAG AATCAATGAA AAGACTTAAC AGCCTCTCA	120
	ATATCCCAGA AAAGTGTCTT GAACAGGGAG GGATGATTIG GAAGATATCT GAAGATAAAC	180
40	AGCTAGCAGT TTGCTGAAA TATGCTGGAG TATTTGCAGA AAATGCAGAA GATGCTGATG	240
	GAAAAGATGT ATTTAATACC AAATCTGTTG GGCTTCTAT TAAAGAGGCA ATGACTTATC	300
	ACCCCAACCA GGTAGTAGAA GGCTGTTGTT CAGATATGGC TGTTACTTTT AATGGACTGA	360
45	CTCCAAATCA GATCCATGTG ATGATGTATG GGGTATACCG CCTTAGGGCA TTTGGCATA	420
	TTTTCAATGA TGCATTGGTT TTCTTACCTC CAAATGGTC TGACAATGAC TGAGAAGTGG	480
50	TAGAAAAGCG TGAATATGAT CTTTGTATAG GACGTGTGTT GTCATTATTT GTAGTAGTAA	540
	CTACATATCC AATACAGCTG TATGTTCTT TTTCCTTTCT AATTGGTGG CACTGGTATA	600
55	ACACACACATT AAAGTCAGTA GTACATTTTT AAATGAGGGT GGTTTTTTC TTTAAACAC	660
	ATGAACATG TAAATGTGTT GGAAAGAAGT GTTTTAAGAA TAATAATTTT GCAAATAAAC	720
	TATTAATAAA TATTTATAATGT GATAAAATCTT AAATTATGAA CATTAGAAAT CTGTGGGCA	780
60	CATATTTTG CTGATTGGTT AAAAAATTAA AACAGGTCTT TAGCGTCTA AGATATGCAA	840

ATGATATCTC TAGTTGTGAA TTTGTGATTA AAGTAAAAC TTTAGCTGTG TGTTCCCTTT	900
ACTTCTGATA CTGATTTATG TTNTAACCG	929

5

10 (2) INFORMATION FOR SEQ ID NO: 99:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 359 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

ATNGGANPCC CCCCNNGCTG CAGGAAATTG CCCGGGCTGC ATGTCTAGTT CCAGTCTGCA	60
CTGGAAAGAA TTCAAATATG CACCTGGCTC CCTTCACTAT TTTGCCCTAT CCTTTGTGCT	120
CATTCTTACT GAAATCTGTC TTGTCAGCTC AGGAATGGGA TTCCCCCAGG AAGGAAAGCA	180
CTTTCTGTT CTGGGAAGCC CAGACTGTTG ACTTTGGGC AGGGACGAAC ATGTGCCTCG	240
TGAATTGCT TGAAAACAGT CACCATCTTC TACCCCCATC ACTGTATAGT GAAAAACCTG	300
ATTAAAGTGG TATCTGAGAA CCAWAAAAAA AAAAAAAAANCTCGAGGG GGGGCCCGG	359

30

35 (2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 952 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

GAATTCCCCG GGGGATCAGG GCAGCCGGGG AGGTGGCCAG GCCAGTGGCA GGCCTGTGGA	60
GACAATCCCT YAGGACTAGG GACAGGGCTG TCCCGGCCTG GGCCACGGCC CACGGACCCG	120
CAGCTCAGGG CCCCTGCCA CGTCGCTGTC CGGGGGTCCG CGCGGGCGT CCCTCGCGTC	180
TCTTCACTGC ACATTGCAAT GCAATTGCGA TTCCCAATTTC TCTGCTAGGA CCCAGCCTGG	240
GTTGGCGCTG CTCCCAGAGC CGGTGGGTCC CAAGANCTTG CGTCCCTTT TGTTCTGTC	300
CCGTTTATCA AGAACACGGG CCCCACCTGT TCACGTTGCC CGAAGGCCAC CCCAAGCCCA	360
ASCCTGCGGG GGGTTCCCM MAYTGCCYTG RAATGCCGG CTTNAAGTTT TTGCGCAACG	420
CMAGGAATTG AGTGTGGGGA CGGCCCCCTGC CGGATTAGGC YTAGCCCTGG CCCAGGTGGT	480
GAGCGGTTTG CAGTGTCCGT TCTCATCCAC CTGATGGGCC CAGATAAAGG CCCCCGCTGT	540

1 —

	CCAGCCTCCC TGGACGGCCC TCGCGGTCCC TGCAGCCAA GATGGGACTC AGACCCGTG	600
5	CCCCAGAGCT CCCCTGCCGC AGAATGGGGC CCCAGCGGC CCCGACCGGG TCCAGGAGCA	660
	CTGCTCGCCT GTACATACGT TTGCCCCTAGC CCACCTGGTG CGGTGGGAGC CACCCCCAGG	720
	TGCNIGGCAC AGCCCCCTCCC CACTCCGCCA CGCCCCCACC CACCCCGCGT GTTTCTGCC	780
10	TGTGACTCCT GGAACCTGCG TCCCTCCCAA AGCCATGGGA GGGGTGTCCCT CCTCAGACCA	840
	TGCCCCCAGA TGATTTTTT AAATAAAAGAA ACAAAATGCAC CTGCAAAAMA AAAAAAAA	900
15	AAAAAAACTC GAGGGGGGGC CCGGTACCCA ATTGCCCTA TAGTGAGCGA TT	952

20 (2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1545 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

30	GAAAGACAAA AGGAAATAGA AGAAAGGGAA AAAAGCGTA AAGACAGACA TGAAGCAAGT	60
	GGGTTTGCAA GGAGACCGAG ATCTCCAACC GGACCTAGCA CGGTGGGCCA CAAGATCATG	120
	CAGAAGTACG GCTTCCGGGA GGGCCAGGGT CTGGGGAAAGC ATGAGCAGGG CCTGAGCACT	180
35	GCCTTGTCAG TGGAGAAGAC CAGCAAGCGT GGCGGCAAGA TCATCGTGGG CGACGCCACA	240
	GAGAAAGGTG TGTCCCCAGG GAAGCGTGTG ACTAGAGGGA AAGGACTGGC CCCATCCATA	300
40	TCAGACATGG CCAGTCTTGA TCCTCATGTG TCACCGAGGG GACAATGAGG CGTGTGGCCA	360
	GAGGGAGAGG GCTGGCCCTG CCATCACTAG AACACAGGCC GTCCCTGTCA TATGATGCAC	420
	TGCCACTTCC GTTTTGTGAA ACCAGGAATC CTGAGGCTCA TCTTTATTTT TTCAGAACAG	480
45	ACGTAGAGAG ATGAAGGCTT GTGGAGGAAA AGATGGTGTAG AGACTTGGC AGAAAATGAG	540
	TAGTCCTCAG GAAGAAATCT TGGTTATGTG TTTAGAGCAT GAAGGACAGA GCCATATAGT	600
50	GTGGCAGTGA ATATACCTGC TATCTCCATC TCAGAGGTGCG TCTCTACTTT TCCCTTTTGC	660
	CCTTTCACTA TAGATGTGAT TTCTGATTCT CTTACAGATT GTTTGCTTTG CGAGATCTGA	720
	TGTTATGTG CAGTCTCTTG GTAAATGATG CCTAGTTGGT GTTTTATTTT CATTAAATT	780
55	TTACAGTCTG TTCTGTGTG AGGGAATTCA GGAAAGAGAC AAACATATGT TAGCATTITA	840
	ATCAGGGAAT TAAGTTTGAG TCAGCCTAGC TGAACCTCCT TTGCTAAAGA AAGAAGAAAA	900
60	CTTTCTGGC AGCCCCGTTA ATGCACAGCT TAGGATAACAT CACGAGCTG ACAGATGCAT	960

	CCAAGAAGTC AGATTCAAAT CCGCTGACTG AAATACTTAA GTGTCTACT AAAGTGGTCT	1020
	TACTAAGGAA CATGGTTGGT GCGGGAGAGG TGGATGAAGA CTTGGGAAGT TGAAACCAAG	1080
5	GAAGAATGTG NAAAAATATG GCAAAGTTGG AAAATGTGTG ATATTGAAA TTCTGGTGC	1140
	CCCTGATGAT GAAGCAGTAC GGATATTTTT AGAATTGAG AGAGTTGAAT CAGCAATTAA	1200
10	AGCGGTTGTT GACTTGAATG GGAGGTATTG TGGTGGACGG GTGGTAAAG CATTTCTA	1260
	CAATTGGAC AAATTCAAGG TCTTGGATTG GGCAGAACAA GTTTGATTAA AAGAACTAGA	1320
	GCACGAGTCA TCTCCGGTGA TCTTAAATG AACTGCAGGC TGAGAAAAGA AGGAAAAGG	1380
15	TCACAGCCTC CATGGCTGTT GCATACCAAG ACTCTTGGAA GGACTTCTAA GATATATGTT	1440
	GATTGATCCC TTTTTTATTT TGTGGTTTTT TAATATAGTA TAAAAATCCT TTTAAAAAAA	1500
20	CAAMAAAAAA AAAAAAAACT CGAGGGGGGG CCCGGTACCC AATTT	1545

25	(2) INFORMATION FOR SEQ ID NO: 102:
	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 1322 base pairs
	(B) TYPE: nucleic acid
30	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:	
35	CTTCTGGAG CGACCGCTCC GCTCGTCTCG TTGGTTCCGG AGGTGCGCTGC GGCGGTGGGA	60
	AATGCTGGCG CGCGCGGCAC GNGGCAGTGG GGCCCTTTTG CTGAGGGGCT CTCTACTGGC	120
	TTCTGGCGCG CTCCTCCGCG CGCCTCCTCT GGATTGCCCG GAAACACCGT GGTACTGTT	180
40	GTGCCGCAGC AGGAGGCCTG GGTGGTGGAG CGAATGGCC GATTCACCG GATCCTGGAG	240
	CCTGGTTGA ACATCCTCAT CCCCCTGTTA GACCGGATCC GATATGTGCA GAGTCTCAAG	300
	GAAATTGTCA TCAACGTGCC TGAGCAGTCG GCTGTGACTC TCGACAATGT AACTCTGCAA	360
45	ATCGATGGAG TCCCTTACCT GCGCATCATG GACCCCTTACA AGGCAAGCTA CGGTGTGGAG	420
	GACCCCTGAGT ATGCCGTAC CCAAGCTAGCT CAAACAAACCA TGAGATCAGA GCTCGGAAA	480
50	CTCTCTCTGG ACAAAAGTCTT CCCGGAACCG GAGTCCTGA ATGCCAGCAT TGTGGATGCC	540
	ATCAACCAAG CTGCTGACTG CTGGGGTATC CGCTGCCTCC GTTATGAGAT CAAGGATATC	600
55	CATGTGCCAC CCCGGGTGAA AGAGTCTATG CAGATGCCAGG TGGAGGGCAGA GCGGGCGAAA	660
	CGGGCCACAG TTCTAGAGTC TGAGGGGACC CGAGAGTCGG CCATCAATGT GGCAGAAGGG	720
	AAGAAACAGG CCCAGATCCT GGCTCCGAA GCAGAAAAGG CTGAACAGAT AAATCAGGCA	780
60	GCAGGAGAGG CCAGTGCAGT TCTGGCGAAG GCGAAGGCTA AAGCTGAAGC TATTGCAATC	840

	CTGGCTGCAG CTCTGACACA ACATAATGGA GATGCACCGAG CTTCACTGAC TGTGCCGAG	900
5	CAGTATGTCA GCGCGTTCTC CAAACTGGCC AAGGACTCCA AACTATCCT ACTGCCCTCC	960
	AACCCCTGGCG ATGTCACCAAG CATGGTGGCT CAGGCCATGG GTGTATATGG AGCCCTCACC	1020
	AAAGCCCCAG TGCCAGGGAC TCCAGACTCA CTCTCCAGTG GGAGCAGCAG AGATGTCCAG	1080
10	GGTACAGATG CAAGTCTTGA TGAGGAACCTT GATCGAGTCA AGATGAGTTA GTGGAGCTGG	1140
	GCTTGGCCAG GGAGTCTGGG GACAAGGAAG CAGATTTTCC TGATTCTGGC TCTAGCTTCC	1200
15	CTGCCAAGAT TTGGTTTTT ATTTCATTTAT TTGAACCTTA GTCGTGTAAT AAACTCACCA	1260
	GTGGCAAACC AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAAAN	1320
	NN	1322

20

(2) INFORMATION FOR SEQ ID NO: 103:

25	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 276 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:	
	NNATAGCTCA ACCATGTTCC AGGAGTGTAT TCCAATCAGC TTGTTTTTC TTAACTGGTT	60
35	AAAGGAATGT TGTCATTCA CCTGCCCAA CTCACATATT AACATTGTT TAACTGGGAT	120
	TAGATAAAAG GAAAGCTGAC TTACAGATGA ACCAAGAGGG AGCTATTATGCCACAGCCC	180
40	CCAGCCCCAGT AACTTTATGT TTCTGATCTC CTGAAAATT TTTTATAAA AAAAGCTTAG	240
	CCAGGAACTA GTAGAAAGAA TAAAGTAAAG ATGGTG	276

45

(2) INFORMATION FOR SEQ ID NO: 104:

50	(i) SEQUENCE CHARACTERISTICS:	
	(A) LENGTH: 381 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
55	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:	
	GATTAAGGTA GAAAAGTACA GAAAACACTA AATTTTCATT GTGCTGTTTC AATGTGGCAG	60
	ATTCTTTAAA ATACTTCGAC ACGCTACAAT AATTAAAGGT TTTAAGAACAA TTAAGATACT	120
60	TAAAAAATAA AAGCCCACAA TTGAATAACA AAAATGAACT TTGTTTTATT TTTTATTGGC	180

5	ATTAATGTAG GTGCCGTGG TGAAAATAGT TTGAAATACT TCACAGTAAC AGTTTGTGC	240
	AGCCCTAGAG ATTAAAAACA GCAAAGTAA TAAGCAGGAC TCTCAACGAC TCATACTCAC	300
	AGACTGTATA ATGTWATCCT ARCACTTCSG GARGCTGARG CGGGAGGATT ACTTGAGCCT	360
	AGGATTGAG ACCAGCCTGG G	381

10

(2) INFORMATION FOR SEQ ID NO: 105:

15	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 638 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
20	(D) TOPOLOGY: linear

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:
--	--------------------------------------------

25	TGTGGAAAAC AGTAGGAAAG CAATGAAAGA AGCTGGTAAG GGAGGCGTCG CTGATTCCAG	60
	AGAGCTAAAG CCGATGGTAG GTGGAGATGA RGARGTGGCC GCCCTCCAAG AATTTCACTT	120
	TCACTTCCTC TCTCTCTCTG TCTTCACTGA CTGCACTTCT TCAGGAGAAG CTTTTGTAT	180
30	CTGTATCACG CAGACATGCT GCTCTTCTG TTGTGTGCT TACCCATCAC TTGGATGGCA	240
	GAATTCTTGT CACAACGTGAG ACACCTYCTA TAAAAGTAAG CTGAAAGGAA CAGCATCCTC	300
	GTCAGTGCTC GGCAGGGCG GGTAGGGAT GATGGTTTT TCCCTAAGGT AAAACTGCTG	360
35	TTGCTCTTGT TTCCCTTTTA ACTGTCAGTG TTGGCTTTC ATCAGACTGA ACATTTGGT	420
	GTACACTTGA ACTGACGGTT TGATTTTAT CATTGGAA GGTGATCATA GCAATTCCCTT	480
40	TCAACTTGCT AAAATTCTATA CTCCCCCTTT TAAAAGTATG GTTCTGCTTA CATTGCTGTC	540
	CTTTTCCCTT GGCTGACTTT TTCTCTGTGTT GCCTAGGTG TACTTTTTTN TTTTTTTNT	600
	TTTCAGTAG CAAACAAGGC TGTTTTCATC AATACCCA	638

45

(2) INFORMATION FOR SEQ ID NO: 106:

50	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 2246 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
55	(D) TOPOLOGY: linear

	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:
--	--------------------------------------------

60	GGCACGGAGGC CGGGGGAGAG TCACGCAAAT GACTTGGAGT GTTCAGGAAA AGGAAAATGC	60
	ACCACGAAGC CGTCAGAGGC AACTTTTCC TGTACCTGTG AGGAGCAGTA CGTGGGTACT	120

	TTCTGTGAAG AATACGATGC TTGCCAGAGG AAACCTTGCC AAAACAACGC GAGCTGTATT	180
5	GATGCCAAATG AAAAGCAAGA TGGGAGCAAT TTCACCTGTG TTTGCCCTTCG TGGTTATACT	240
	GGAGAGCTTT GCCAGTCCAA GATTGATTAC TGCACTCTAG ACCCATGCAG AAATGGAGCA	300
	ACATGCAATT CCAGTCTCAG TGGATTCAAC TGCCAGTGTG CAGAAGGATA CTTGGATCT	360
10	GCTTGTGAAG AAAAGGTGGA CCCCTGCGCC TCGTCTCCGT GCCAGAACAA CGGCACCTGC	420
	TATGTGGACG GGGTACACTT TACCTGCAAC TGCAGCCGG CCTTCACAGG GCCGACCTGT	480
15	GCCCAGCTTA TTGACTTCTG TGCCCTCAGC CCCTGTGCTC ATGGCACGTG CCGCAGCGTG	540
	GGCACCAGCT ACAAAATGCCT CTGTGATCCA GGTTACCATG GCCTCTACTG TGAGGAGGAA	600
	TATAATGAGT GCCTCTCCGC TCCATGCCTG AATGCAGCCA CCTGCAGGGA CCTCGTTAAT	660
20	GGCTATGAGT GTGCTGTGCCT GCCAGAAATAC AAAGGAACAC ACTGTGAATT GTACAAGGAT	720
	CCCTGCGCTA ACAGTCAGCTG TCTGAACGGA GCCACCTGTG ACAGCGACGG CCTGAATGGC	780
25	ACGTGCATCT GTGCCACCCGG GTTTACAGGT GAAGAGTGGC ACATTGACAT AAAATGAATGT	840
	GACAGTAACC CCTGCCACCA TGGTGGGACC TGCCTGGACC AGCCCAATGG TTATAACTGC	900
	CACTGCCCCG ATGGTTGGGT GGGAGCAAAC TGTGAGATCC ACCTCCAATG GAAGTCCGGG	960
30	CACATGGCGG AGAGCCTCAC CAACATGCCA CGGCACCTCCC TCTACATCAT CATTGGAGCC	1020
	CTCTGCGTGG CCTTCATCCT TATGCTGATC ATCCCTGATCG TGGGGATTTG CCGCATCAGC	1080
35	CGCATTGAAT ACCAGGGTTC TTCCAGGCCA GCCTATGAGG AGTTCTACAA CTGCCGCAGC	1140
	ATCGACAGCG AGTTTCAGCAA TGCCATTGCA TCCATCCGGC ATGCCAGGTT TGGAAAGAAA	1200
	TCCCCGGCTG CAATGTATGA TGTGAGCCCC ATCCCTATG AAGATTACAG TCCTGATGAC	1260
40	AAACCCCTGG TCACACTGAT TAAAACAAA GATTTGTAAT CTTTTTTTGG ATTATTTTC	1320
	AAAAAGATGA GATACTACAC TCATTTAAAT ATTTTTAAGG AAAWAAAAA GCTTAAGAAA	1380
45	TTTAAATGC TAGCTGCTCA AGRGTTTCA GTAGAAATATT TAAGAACTAA TTTCTGCAG	1440
	CTTTAGTTT GGAAAAAAATA TTTTAAAAAC AAAATTTGTG AAACCTATAG ACGATGTTTT	1500
	AATGTACCTT CAGCTCTCTA AACTGTGTGC TTCTACTAGT GTGTGCTCTT TTCACTGTAG	1560
50	ACACTATCAC GAGACCCAGA TTAATTTCTG TGGTTGTTAC AGAATAAGTC TAATCAAGGA	1620
	GAAGTTTCTG TTTGACGTTT GAGTGCCTGGC TTCTGAGTA GAGTTAGGAA AACACCGTAA	1680
55	CGTAGCATAT GATGTATAAT AGAGTATAACC CGTTACTTAA AAAGAAGTCT GAAATGTTCG	1740
	TTTTGTGGAA AAGAAACTAG TTAAATTTAC TATTCCTAAC CCGAATGAAA TTAGCCTTGT	1800
	CCTTATTCCTG TGCACTGGGTA AGTAACCTTAT TTCTGCACTG TTTTGTGAA CTTTGTGGAA	1860
60	ACATCTTTC GACTTGTGTT TTGTCATTTT CGTAACAGTC GTCGAACTAG GCCTCAAAAA	1920

	CATACTAAC GAAAAGGCCT AGCGAGGCAA ATTCTGATTG ATTTGAATCT ATATTTTCT	1980
5	TTAAAAAGTC AAGGGTTCTA TATTGTGAGT AAATTAATT TACATTGAG TTGTTTGTG	2040
	CTAAGAGGTA GTAAATGTAA GAGAGTACTG GTTCCTTCAG TAGTGAGTAT TTCTCATAGT	2100
	GCAGCTTTAT TTATCTCCAG GATGTTTTG TGGCTGTATT TGATTGATAT GTGCTTCCTTC	2160
10	TGATTCTTGC TAATTTCCAA CCATATTGAA TAAATGTGAT CAAGTCAAAA AAAAAAAA	2220
	AAAAAAAATT ACTCGGTGCG AAGGGA	2246

15

(2) INFORMATION FOR SEQ ID NO: 107:

	(i) SEQUENCE CHARACTERISTICS:
20	(A) LENGTH: 1105 base pairs
	(B) TYPE: nucleic acid
	(C) STRANDEDNESS: double
	(D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

	GAATTGGCA GAGCCCACCTT AGAGGAGCTA AAATAGCTAA AGGTTACATG CTTTGCCTCA	60
30	AATAATAGAC TTAGTGAAGA GGGTAGAAGT AGAAATRAGG TCAGCCCCC AGAGCAGTCT	120
	GGTGGCCCTT AGCAACCAGG AAGGTAAAGC CGGTACCTCA GTTAAATCAC CAAGTTACT	180
	GGAAGTGCAT ATTTTCATG TCCAAATTC AGTAAGTCAT GGACCAAATG TTTATTTG	240
35	TATGCTTAA AAAGTTGCTT GCTTCTTGTA AGTTTCTCA GTGGAAGGGT TCCAAGTTAT	300
	GACTTAATCT ATGTTTGCAG CATTGCACTG GAAACAGGAT TTGTCGTGA AATGGCTCTG	360
40	TCATTTCGTC ACCACTCTG TAGGGAGATT GTGGATTAG GAAGGCCAGA AGCAACAGCA	420
	GATATGCCCTG GTGTTGAAT GGATGTGCCT CTYTCGGAGG CACCAAGCAG CATAACCATA	480
	TTATAAAGTT TTGATTTC TAACATCTGA AGACAGGCAT CCAGCCTTGC AGAACAGCCA	540
45	GGTGTCTGTT CTATAGACTA CAGTTCCCTG TTTCCAGAAT TACGGTAACC AAATAATACA	600
	CAAGGTCACT TGAAITGCACT TCCCAACAAC CTGAACAAAG AGCACCTTGT CGCTTGCTGG	660
	TAGGTGCTGT ACCAGACTCT TTGTAATCTG CCTTAGKTCA GRGAAGAACAA AGCCATTACC	720
50	AGTATGGGAG TCCATCCYTA GTCAGGGCTA GTTGCTATTA TCCCTTGAAT ACTCTGCAGG	780
	CATCCCACAA GACATTGAG ACTTCATAATT TGTCAAATAA TAGAAATSTG GCTGGCTAG	840
55	TGGCTCATGC CTGTAATCCT AACCCCTTG GAGGCTGATG TGGGCAGATT GCTTGAGGCC	900
	AGGAGTTGAG GACCCACCTG GGCAACACAG TGACATGTTG TCTCTACAAA AAATTTAAA	960
60	ATTAACCTAGG CATGGTAGTG TGCCTATAGT CCCAGCTACT CCAGAGGCTG AGGCAGGAAG	1020

ATCCCTTGAG CCCAGTAATT CAAGGCTACA GTTAGCTCTG ATCCTGCCAC TGCACCTCG	1080
TCTTGGTAAA GGAGCTAAAC CCAGT	1105

5

(2) INFORMATION FOR SEQ ID NO: 108:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 505 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

ATTCACACA GGAAACAGCT ATGACCATGA TTCCGCAAG CNCGAAATTA ACCNTCACTA	60
20 AAGGAAACAA AACTGGAGCT CCACCGCGGT GGCGGCCGCT CTAGAACTAG TGGATCCCC	120
GGCCTCAGGA ATTCCGCACG AGTTCTTCCA CATGTGTGCA CCCCCAGCTT GGCCAACCC	180
25 CAGCCTTGGG GTGGGGCCCG AAGCATCTTC CCTTCCGCTT GGCGTCCTG GGATGGAT	240
GAGTGCCTGG CTCCCCATCTC CTCCCTCACCT TTTGTTGCTA TCGGCAGCTG CTGGCTCAGG	300
GGCATCCCCAC CTCCGGGCTC TGGTTCCCTC TGCCCTGGAA GGGCTCCAGG ACCCGTCCCA	360
30 ATAACCACCC ACGGCCAGGA GRGCCAAGGC CCCGTGCTGG ATATTAAAT TTAGGGCCG	420
GTCTCCAGGG CGCGTAGATA AATAAATACA CTCAGCGTCA AAAAAAAA AAAAAAAA	480
35 AAAAAAAA AAAAAAAA CTCGA	505

35

(2) INFORMATION FOR SEQ ID NO: 109:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1380 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

50 AATCATGAGC CTCCAGAAGA GACAGATGGC CCACCAAGGAG CTGTTGCTCT GGTTGCCTTC	60
CTGCAGGCCT TGGAGAAGGA GGTCGCCATA ATCGTTGACC AGAGAGCCTG GAACTTGCAC	120
CARAAGATTG TTGAAGATGC TGTGAGCAA GGTGTTCTGA AGACGCAGAT CCCGATATTA	180
55 ACTTACCAAG GTGGATCACT GGAAGCTGCT CAGGCATTCC TGTGAAAAA TGGGGACCCG	240
CAGACACCTA GATTTGACCA CCTGGTGGCC ATAGAGCGTG CGGAAGAGC TGCTGATGGC	300
AATTACTACA ATGCAAGGAA GATGAACATC AAGCACTTGG TTGACCCAT TGACGATCTT	360

60

	TTTCTTGCTG CGAAGAAGAT TCCCTGGAATC TCATCAACTG GAGTCGGTGA TGGAGGCAAC	420
	GAGCTTGGGA TGGGTAAGT CAAGGAGGCT GTGAGGAGGC ACATAACGCCA CGGGRATGTC	480
5	ATCCCTGCG ACCTGGAGGC TGACTTTGCC GTCATTGCTG GTGTTCTAA CTGGGGAGGC	540
	TATGCCCTGG CCTGCGCACT CTACATCCTG TACTCATGTG CTGTCCACAG TCAGTACCTG	600
10	AGGAAAGCAG TCGGACCCTC CAGGGCACCT GGAGATCAGG CCTGGACTCA GGCCCTCCCG	660
	TCGGTCATTA AGGAAGAAAA AATGCTGGGC ATCTTGGTGC AGCACAAAGT CCGGAGTGGC	720
	GTCTCGGGCA TCGTGGGCAT GGAGGTGGAT GGGCTGCCCT TCCACAAACAC CCACGCCAG	780
15	ATGATCCAGA AGCTGGTGGA CGTCACCACCG GCACAGGTGT AACCGTCCAT GTTCCGTGTG	840
	AGCAGAGTCC CTACCAACGG GCAGGTCTGC ATCCGGGAG AATGCAGCTG CTTCTGGCGA	900
20	CAATCCTGCT AGTAAACACT GGTCTTCGGT GAGCAACGAA CACTCGCCTG GCCTGGAAA	960
	CTGCATGCC ACTTTCTGGG AGGGGTTAGT GCAGGTGCCG TGGACAAAGG ACAACATTTC	1020
	TCTGGGGCTT TTAAACTTTT ATTCTTAAGA CTCTAAAGGC GTTGATTICA ACCCTCCTTC	1080
25	ACTCTGGCTT CTTCAGGCAA CCCACGTGGT CTCCCTGTGAG AATCTTCTCG ACAGTTACTT	1140
	ATGGGGACAC TTGTGAACAA TTAACTGCCA GGCAGAGCAT GAGAACAAAC ATTCCCAGGC	1200
30	CATGAGGAT AGGATACTCC AGACTCCAGT CATCCTCCCC CATCCATGGT TTCTGTTACT	1260
	CATGGTTTCA GTTACTCATA GCCAACTGCA GACCGAAAAT ACTAAATGAA AAATTCAGA	1320
	AATAAAACAAC TCTTAAGTTT TAAAAAAAAA AAAAAAAAAA AAAAAAAAAA GGGCGGCCGC	1380
35		

(2) INFORMATION FOR SEQ ID NO: 110:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 646 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

	CAGATGCCAG GGACTTGGNC TTCCCCCGGT TGAACCACAG GTTCCAAGAA ACCTGCAGGG	60
50	TCCAGCCTCC CCCCCATCCC CAGTYTTCCC CACCCCTGGCC CGGCCCTCCA GGTGCAGAAA	120
	CATGCAGGCC CCTCTCCAGG ACTGTGGGAG GAGTGTGTCC CTCAGACTGG CCTGTGTCCCT	180
55	GGCTCCCTTT ACCACCTCTT CCAGAGGTG TCACCTGCAG CTGCCCCAGG ATAAAGGCAA	240
	GGCCAGARAG GACTCCTGAA CTCCCTGTGTG CCTGGGGTGG CAGGGGCAA CATGCCAAC	300
	TGGTGGCCTG AGCGGGGCCA TGGTGARGAC ACCCTTGGTG GCTTGTCCCA CATCAAGCTG	360
60	GGGARGTGACA CTTAGGATGC ATTTTCAAT ATTTTAGTGT TTGAATAACG GGCTAWCTTG	420

AGAAAAAAAT AATTTGAATC ACACATCACA CCAAAAATAA ATTCTAGGTG GATTTTAACA 480
 5 CTTTCCAAA ATTATTATTA GTTTAGAGAC AGGGTCTCAC TCCGTGGCT AGGCTGGAGT 540
 GCANGGTAT GATCATGGTT CACTGCAACC TTAAACTCCC TGGCCTCATA TGATCCCCC 600
 GGGCTCCAGC CCCTCCAAAG TTACTGGGAA ACTACCAAAC ATGCC 646

10

(2) INFORMATION FOR SEQ ID NO: 111:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

20

Met	Asp	Ser	Tyr	Trp	His	Ser	Arg	Cys	Leu	Lys	Cys	Ser	Cys	Cys	Gln
1					5				10					15	

Ala	Xaa	Trp	Ala	Thr	Ser	Ala	Arg	Pro	Val	Thr	Pro	Lys	Val	Ala	Xaa
					20				25				30		

30

(2) INFORMATION FOR SEQ ID NO: 112:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

40

Ile	Tyr	Ser	Ser	Gly	Tyr	Phe	Gln	Ile	Tyr	Asn	Met	Leu	Leu	Leu	Thr
1					5				10					15	

Ile	Leu	Ile	Leu	Leu	Cys	Asn	Arg	Thr	Pro	Glu	Leu	Ile	Pro	Gly	Phe
					20			25					30		

Tyr	Ile	Arg	Xaa												
			35												

50

(2) INFORMATION FOR SEQ ID NO: 113:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 220 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

60

Met	Ser	His	Lys	Leu	Gly	Asp	Pro	Gly	Phe	Val	Val	Phe	Ala	Thr	Leu
1					5				10				15		

	Val	Val	Ile	Val	Ala	Leu	Ile	Leu	Ile	Phe	Val	Val	Gly	Pro	Arg	His
	20						25					30				
5	Gly	Gln	Thr	Asn	Ile	Leu	Val	Tyr	Ile	Thr	Ile	Cys	Ser	Val	Ile	Gly
	35						40					45				
	Ala	Phe	Ser	Val	Ser	Cys	Val	Lys	Gly	Leu	Gly	Ile	Ala	Ile	Lys	Glu
10	50					55				60						
	Leu	Phe	Ala	Gly	Lys	Pro	Val	Leu	Arg	His	Pro	Leu	Ala	Trp	Ile	Leu
	65					70			75			80				
15	Leu	Leu	Ser	Leu	Ile	Val	Cys	Val	Ser	Thr	Gln	Ile	Asn	Tyr	Leu	Asn
	85					90					95					
	Arg	Ala	Leu	Asp	Ile	Phe	Asn	Thr	Ser	Ile	Val	Thr	Pro	Ile	Tyr	Tyr
	100					105					110					
20	Val	Phe	Phe	Thr	Thr	Ser	Val	Leu	Thr	Cys	Ser	Ala	Ile	Leu	Phe	Lys
	115					120					125					
	Glu	Trp	Gln	Asp	Met	Pro	Val	Asp	Asp	Val	Ile	Gly	Thr	Leu	Ser	Gly
25	130					135					140					
	Phe	Phe	Thr	Ile	Ile	Val	Gly	Ile	Phe	Leu	Leu	His	Ala	Phe	Lys	Asp
	145					150				155		160				
30	Val	Ser	Phe	Ser	Leu	Ala	Ser	Leu	Pro	Val	Ser	Phe	Arg	Lys	Asp	Glu
	165					170				175						
	Lys	Ala	Met	Asn	Gly	Asn	Leu	Ser	Asn	Met	Tyr	Glu	Val	Leu	Asn	Asn
	180					185				190						
35	Asn	Glu	Glu	Ser	Leu	Thr	Cys	Gly	Ile	Glu	Gln	His	Thr	Gly	Glu	Asn
	195					200				205						
	Val	Ser	Arg	Arg	Asn	Gly	Asn	Leu	Thr	Ala	Phe	Xaa				
40	210					215				220						

(2) INFORMATION FOR SEQ ID NO: 114:

45	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 32 amino acids
	(B) TYPE: amino acid
	(D) TOPOLOGY: linear
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:
	Met Thr Ile Trp Glu Arg Lys Tyr Ile Trp Met Leu Gln Ile Cys Val
	1 5 10 15
55	Phe Leu Glu Pro Arg Ala Lys Pro Ser Leu Gly Asp Leu Asp Trp Xaa
	20 25 30

(2) INFORMATION FOR SEQ ID NO: 115:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 27 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

10 Met Leu Thr Phe Leu Leu Phe Ile Pro Val Ala Pro Thr Glu Thr Ser
 1 5 10 15

Gln Lys Asn Arg Ser Val Phe Leu Pro Pro Xaa
 20 25

15

(2) INFORMATION FOR SEQ ID NO: 116:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 132 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

25 Met Leu Phe Val Phe Cys Cys Thr Val Phe Phe Val Cys Leu Phe Val
 1 5 10 15

30 Tyr Leu Val Gly Phe Leu Glu Arg Glu Ile Trp Lys Arg Asp Ile His
 20 25 30

Lys Ser Tyr Thr Pro Thr Phe Pro Phe Tyr His Asp Ile Gln Glu Glu
 35 40 45

35 Thr Ser Arg Ala Lys Asn Gly Val Lys Lys Gly Ser Met Ala Gly Thr
 50 55 60

Ser Lys Glu Leu Arg Ala Val Ala Leu Lys Asn Tyr Phe Phe Tyr Tyr
 65 70 75 80

40 Tyr Phe Glu Ser Met Glu Val Phe His Ser Leu Gly Lys Gly Gly Lys
 85 90 95

45 Ser Ala Phe Ile Phe Ile Gln Ser Tyr Leu Ile Thr Ser Lys Thr His
 100 105 110

Met Leu Glu Ile Ala Phe Ala Gly Ala Lys Tyr Ile Asn Glu Gln Glu
 115 120 125

50 Tyr Ile His Xaa
 130

55 (2) INFORMATION FOR SEQ ID NO: 117:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 65 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

Met Trp Tyr Phe Met Ser Leu Ile Ser Met Val Leu Leu Ser Pro
1 5 10 15

5 Ser Cys Ser Asp Leu Leu Val Ile Ser Val Leu Asn Leu Glu Gln Arg
20 25 30

10 Arg Gln Ser Lys Val Gly Phe Glu Pro Phe Thr Ser Pro Leu Cys Gly
35 40 45

Xaa Trp His His Leu Ser Pro Asp Arg Leu Pro Gln Asp Gly Thr Phe
50 55 60

15 Xaa
65

20 (2) INFORMATION FOR SEQ ID NO: 118:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

Leu Leu Leu Phe Cys Ile Leu Gly Xaa
1 5

30 (2) INFORMATION FOR SEQ ID NO: 119:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

40 Met Gly Val Leu Phe Val Pro Gln Glu Thr Ser Xaa Lys Val Xaa Xaa
1 5 10 15

45 Asp Ile Xaa Gly Leu Ser Gln Phe Val Met Gly Glu Lys Arg Thr Thr
20 25 30

Ser Ile Arg Gly Ile Gln Ala Arg Tyr Gln Val Asp Arg Gly Leu Glu
35 40 45

50 Tyr Cys
50

55 (2) INFORMATION FOR SEQ ID NO: 120:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 76 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

5 Met Leu Trp Thr Cys Gln
 1 5 10 15
Lys Ala Leu Val Arg Arg Gln Phe Cys Leu Phe Asn Leu Ile Ala Arg
 20 25 30
10 Asn Ser Ser Leu Met Leu Gln Lys Asp Glu Lys Lys Gly Lys Lys Arg
 35 40 45
Asp Asn Ser Gln Ala Gln Arg Glu Lys Lys Gly Gly Lys Glu Pro
 50 55 60
15 Gln Gly Asp Leu Gln Glu Arg Pro Gly Pro Gly Xaa
 65 70 75

20 (2) INFORMATION FOR SEQ ID NO: 121:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

25 Met His Asn Ala Phe Asn Leu Asn Val Leu Thr Leu Phe Leu Ser Val
 1 5 10 15
30 Leu Cys Cys Thr Phe Ser Asp Ser Glu Leu Xaa
 20 25

35

(2) INFORMATION FOR SEQ ID NO: 122:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 122:

40 Met Ser Trp Leu Phe Leu Leu Phe Ala Leu Leu Cys Lys Phe Gln His
 1 5 10 15
45 Lys Leu Xaa Phe His Asn Ile Xaa
 20

50

(2) INFORMATION FOR SEQ ID NO: 123:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 123:

60 Met Leu Leu Phe Leu Thr Val Ile Asn Phe Met Ala Leu Ala Lys Met

1 5 10 15

Asn Phe Cys Gly Asp Xaa
20

5

(2) INFORMATION FOR SEQ ID NO: 124:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

15 Met Val Xaa Asn Leu Gln Val Ile Ser Ile Trp Xaa Xaa Ser Thr Thr
1 5 10 15

20 Cys Phe Tyr Ala Cys Ile Trp Xaa Gln Gly Cys Leu Met Leu Arg Xaa
20 25 30

Phe Xaa Thr Leu Asn Asn Val Thr Arg Leu Pro Ser Ser Gln Lys Pro
35 40 45

25 Ile Lys Cys Tyr Leu Leu Xaa
50 55

30 (2) INFORMATION FOR SEQ ID NO: 125:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 318 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

35 Met Leu Ser Glu Ser Ser Phe Leu Lys Gly Val Met Leu Gly Ser
1 5 10 15

40 Ile Phe Cys Ala Leu Ile Thr Met Leu Gly His Ile Arg Ile Gly His
20 25 30

45 Gly Asn Arg Met His His His Glu His His His Leu Gln Ala Pro Asn
35 40 45

Lys Glu Asp Ile Leu Lys Ile Ser Glu Asp Glu Arg Met Glu Leu Ser
50 55 60

50 Lys Ser Phe Arg Val Tyr Cys Ile Ile Leu Val Lys Pro Lys Asp Val
65 70 75 80

55 Ser Leu Trp Ala Ala Val Lys Glu Thr Trp Thr Lys His Cys Asp Lys
85 90 95

Ala Glu Phe Phe Ser Ser Glu Asn Val Lys Val Phe Glu Ser Ile Asn
100 105 110

60 Met Asp Thr Asn Asp Met Trp Leu Met Met Arg Lys Ala Tyr Lys Tyr
115 120 125

Ala Phe Xaa Lys Tyr Arg Asp Gln Tyr Asn Trp Phe Phe Leu Ala Arg
 130 135 140

5 Pro Thr Thr Phe Ala Ile Ile Glu Asn Leu Lys Tyr Phe Leu Leu Lys
 145 150 155 160

Lys Asp Pro Ser Gln Pro Phe Tyr Leu Gly His Thr Ile Lys Ser Gly
 165 170 175

10 Asp Leu Glu Tyr Val Gly Met Glu Gly Ile Val Leu Ser Val Glu
 180 185 190

15 Ser Met Lys Arg Leu Asn Ser Leu Leu Asn Ile Pro Glu Lys Cys Pro
 195 200 205

Glu Gln Gly Gly Met Ile Trp Lys Ile Ser Glu Asp Lys Gln Leu Ala
 210 215 220

20 Val Cys Leu Lys Tyr Ala Gly Val Phe Ala Glu Asn Ala Glu Asp Ala
 225 230 235 240

Asp Gly Lys Asp Val Phe Asn Thr Lys Ser Val Gly Leu Ser Ile Lys
 245 250 255

25 Glu Ala Met Thr Tyr His Pro Asn Gln Val Val Glu Gly Cys Cys Ser
 260 265 270

30 Asp Met Ala Val Thr Phe Asn Gly Leu Thr Pro Asn Gln Met His Val
 275 280 285

Met Met Tyr Gly Val Tyr Arg Leu Arg Ala Phe Gly His Ile Phe Asn
 290 295 300

35 Asp Ala Leu Val Phe Leu Pro Pro Asn Gly Ser Asp Asn Asp
 305 310 315

40 (2) INFORMATION FOR SEQ ID NO: 126:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

Met Thr Trp Pro Pro Ser Cys Leu Val Ala Leu Leu Leu Ser Thr Val
 1 5 10 15

50 Thr Gln Lys Met Thr Pro Leu Asn Leu Met Arg Thr Thr Gly Pro Ile
 20 25 30

55 Asn Ser Phe Cys Leu Leu Pro Thr Phe Phe Phe Pro Ser Tyr Leu
 35 40 45

Pro Ser Leu Met Pro Thr Pro Thr Asp Pro Xaa
 50 55

(2) INFORMATION FOR SEQ ID NO: 127:

5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 99 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

10 Ile Leu Phe Ser Phe Leu Ile Pro Ser Asn Leu Ser Phe Ser Pro Val
 1 5 10 15

15 Ile Phe Phe Leu Cys Gly Pro Phe Lys Val Val Ile Ile Cys Thr Glu
 20 25 30

15 Leu Gln Asn Val Ser Arg Ser Pro Gln Thr Thr Leu Ala Thr Val Tyr
 35 40 45

20 Cys Asn Lys Ile Thr Ser Tyr Ile Cys Arg Asn Ser Phe Gly Val Ile
 50 55 60

Leu Phe Phe Pro Leu Asn Ile Tyr Asn Trp Thr Asn Ala Gly Lys Lys
 65 70 75 80

25 Lys Lys Met Val Ser Lys Lys Pro Lys Ile Lys Phe Arg Gly His Gln
 85 90 95

30 Ala Phe Xaa

(2) INFORMATION FOR SEQ ID NO: 128:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

40 Met Ser Ile Leu Leu Leu Xaa Phe Pro Ser Ala Pro Ala Pro Val Val
 1 5 10 15

45 Ser Gly Gly Leu Gln Pro Trp Leu His Ser Cys Ile Xaa
 20 25

(2) INFORMATION FOR SEQ ID NO: 129:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

55 Met Gly Thr Ser Leu Asn Leu Gln Ile Met Ala Leu Phe Ser Gly Gln
 1 5 10 15

60 Ala Met Ala Pro Arg Xaa

20

5 (2) INFORMATION FOR SEQ ID NO: 130:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 112 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

Met Leu Trp Leu Pro Leu Leu Ala Ala Leu Ser Pro Ser Pro Pro Gly
1 5 10 15

15 Val Ser Ser Glu Glu Glu Gln His Trp Ser Gln Ala Glu Ala Leu Pro
20 25 30

20 Cys Trp Asp Pro Gly Ser Glu Ser Ser Pro Arg Ile Pro Gly Cys Arg
35 40 45

Glu Leu Gln Ser Cys Pro Pro Pro Thr Ala Pro Ser Ala His Thr Gln
50 55 60

25 Ser Pro Gly Gly Leu Gly Ala Lys Ala Gly Ala Ala Leu Val Pro Phe
65 70 75 80

30 Pro Gly Pro Ser Phe Pro Thr Ser Lys Pro Lys Lys Gly Glu Ala Gly
85 90 95

Ala Pro Val Pro Gln Pro His Ser Ala Leu Thr Val Pro Ser Ser Xaa
100 105 110

35

40 (2) INFORMATION FOR SEQ ID NO: 131:
40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

Met Glu Lys Pro Leu Phe Pro Leu Val Pro Leu His Trp Phe Gly Phe
1 5 10 15

50 Gly Tyr Thr Ala Leu Val Val Ser Gly Gly Ile Val Gly Tyr Val Lys
20 25 30

55 Thr Gly Ser Val Pro Ser Leu Ala Ala Gly Leu Leu Phe Gly Ser Leu
35 40 45

Ala Gly Leu Gly Ala Tyr Gln Leu Tyr Gln Asp Pro Arg Asn Val Trp
50 55 60

60 Gly Phe Leu Ala Ala Thr Ser Val Thr Phe Val Gly Val Met Gly Met
65 70 75 80

Arg Ser Tyr Tyr Gly Lys Phe Met Pro Val Gly Leu Ile Ala Gly
85 90 95

5 Ala Ser Leu Leu Met Ala Ala Lys Val Gly Val Arg Met Leu Met Thr
100 105 110

Ser Asp

10

(2) INFORMATION FOR SEQ ID NO: 132:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

20 Met Ile Thr Leu Leu Ile Trp Met Leu Ala Gly Phe Ile Ala Arg Ile
1 5 10 15

25 Xaa Val Ala Leu Gln Xaa
20

(2) INFORMATION FOR SEQ ID NO: 133:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

Met Ala Gly Val Ser Glu Ile Ser Val Cys Phe Xaa Leu Leu Ser Leu
1 5 10 15

40 Phe Ser Leu Phe Cys Ser Phe Tyr Phe Pro Lys Gln Ala Thr Pro Lys
20 25 30

45 Arg Asp Leu Phe Val Gln Glu Ser Gly Lys Gly Lys Arg Asn Thr Glu
35 40 45

45 Ser Trp Glu Xaa
50

50

(2) INFORMATION FOR SEQ ID NO: 134:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 99 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

60 Met Thr Ser Ala Leu Thr Gln Gly Leu Glu Arg Ile Pro Asp Gln Leu
1 5 10 15

Gly Tyr Leu Val Leu Ser Glu Gly Ala Val Leu Ala Ser Ser Gly Asp
 20 25 30

5 Leu Glu Asn Asp Glu Glu Gln Ala Ala Ser Ala Ile Ser Glu Leu Val Ser
 35 40 45

Thr Ala Cys Gly Phe Arg Leu His Arg Gly Met Asn Val Pro Phe Lys
 10 50 55 60

Arg Leu Ser Val Val Phe Gly Glu His Thr Leu Leu Val Thr Val Ser
 65 70 75 80

15 Gly Gln Arg Val Phe Val Val Lys Arg Gln Asn Arg Gly Arg Glu Pro
 85 90 95

Ile Asp Val

20

(2) INFORMATION FOR SEQ ID NO: 135:

(i) SEQUENCE CHARACTERISTICS:
 25 (A) LENGTH: 176 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

30 Met Gly Ser Ala Ala Leu Glu Ile Leu Gly Leu Val Leu Cys Leu Val
 1 5 10 15

Gly Trp Gly Gly Leu Ile Leu Ala Cys Gly Leu Pro Met Trp Gln Val
 35 20 25 30

Thr Ala Phe Leu Asp His Asn Ile Val Thr Ala Gln Thr Thr Trp Lys
 35 40 45

40 Gly Leu Trp Met Ser Cys Val Val Gln Ser Thr Gly His Met Gln Cys
 50 55 60

Lys Val Tyr Asp Ser Val Leu Ala Leu Ser Thr Glu Val Gln Ala Ala
 65 70 75 80

45 Arg Ala Leu Thr Val Ser Ala Val Leu Leu Ala Phe Val Ala Leu Phe
 85 90 95

Val Thr Leu Ala Gly Ala Gln Cys Thr Thr Cys Val Ala Pro Gly Pro
 50 100 105 110

Ala Lys Ala Arg Val Ala Leu Thr Gly Gly Val Leu Tyr Leu Phe Cys
 115 120 125

55 Gly Leu Leu Ala Leu Val Pro Leu Cys Trp Phe Ala Asn Ile Val Val
 130 135 140

Arg Glu Phe Tyr Asp Pro Ser Val Pro Val Ser Gln Lys Tyr Glu Leu
 145 150 155 160

60 Gly Ala Xaa Cys Thr Ser Ala Gly Arg Pro Pro Arg Cys Ser Trp Xaa

165

170

175

5

(2) INFORMATION FOR SEQ ID NO: 136:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 187 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

Met	Val	Leu	Leu	Trp	Val	Val	Thr	Cys	Pro	Ala	Thr	Met	Leu	Thr	Glu
1	5													15	

Pro	Gln	Asn	Pro	His	Leu	Ile	Gly	Phe	Val	Ala	Tyr	Ser	Gly	Pro	Ser
	20													30	

His	Thr	Thr	Gln	Pro	His	Lys	Tyr	Trp	Leu	Leu	Leu	Asp	Gly	Gln	Ala
	35													45	

Asp	Pro	Ala	Ala	Ala	Glu	Gly	Pro	Val	Lys	Arg	Lys	Ala	Ala	Ser	Val
	50													60	

Val	Trp	Trp	Pro	Gln	Ala	Leu	Arg	His	Leu	Ser	Leu	Leu	Val	His	Cys
	65													80	

Trp	Glu	Glu	Ser	Tyr	Glu	Met	Asn	Ile	Gly	Cys	Gln	Ser	Leu	Trp	Ala
	85													95	

Gly	Gly	Leu	Ala	Ser	Ser	Gly	Asn	Gly	Trp	Asp	Leu	Gly	Val	Ala	Phe
	35													110	

Arg	Arg	Asp	Thr	Cys	Met	Ser	Ser	Ser	Leu	His	Trp	Lys	Glu	Phe	
	115													125	

Lys	Tyr	Ala	Pro	Gly	Ser	Leu	His	Tyr	Phe	Ala	Leu	Ser	Phe	Val	Leu
	40													140	

Ile	Leu	Thr	Glu	Ile	Cys	Leu	Val	Ser	Ser	Gly	Met	Gly	Phe	Pro	Gln
	145													160	

Glu	Gly	Lys	His	Phe	Ser	Val	Leu	Gly	Ser	Pro	Asp	Cys	Ser	Leu	Trp
	45													175	

Gly	Arg	Asp	Glu	His	Val	Pro	Arg	Glu	Phe	Ala					
	50													185	

(2) INFORMATION FOR SEQ ID NO: 137:

55 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 288 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

Met Pro Ala His Arg Phe Val Leu Ala Val Gly Ser Ala Val Phe Asn
 1 5 10 15

5 Ala Met Phe Asn Gly Gly Met Ala Thr Thr Ser Thr Glu Ile Glu Leu
 20 25 30

Pro Asp Val Glu Pro Ala Ala Phe Leu Ala Leu Leu Lys Phe Leu Tyr
 35 40 45

10 Ser Asp Glu Val Gln Ile Gly Pro Glu Thr Val Met Thr Thr Xaa Tyr
 50 55 60

Thr Ala Lys Lys Tyr Ala Val Pro Ala Leu Glu Ala His Cys Val Glu
 15 65 70 75 80

Phe Leu Lys Lys Asn Leu Arg Ala Asp Asn Ala Phe Met Leu Leu Thr
 85 90 95

20 Gln Ala Arg Leu Phe Asp Glu Pro Gln Leu Ala Ser Leu Cys Leu Glu
 100 105 110

Asn Ile Asp Lys Asn Thr Ala Asp Ala Ile Thr Ala Glu Gly Phe Thr
 115 120 125

25 Asp Ile Asp Leu Asp Thr Leu Val Ala Val Leu Glu Arg Asp Thr Leu
 130 135 140

Gly Ile Arg Glu Val Arg Leu Phe Asn Ala Val Val Arg Trp Ser Glu
 30 145 150 155 160

Ala Glu Cys Gln Arg Gln Gln Leu Gln Val Thr Pro Glu Asn Arg Arg
 165 170 175

35 Lys Val Leu Gly Lys Ala Leu Gly Leu Ile Arg Phe Pro Leu Met Thr
 180 185 190

Ile Glu Glu Phe Ala Ala Gly Pro Ala Gln Ser Gly Ile Leu Val Asp
 40 195 200 205

Arg Glu Val Val Ser Leu Phe Cys Thr Ser Pro Ser Thr Pro Ser His
 210 215 220

45 Glu Trp Ser Ser Leu Thr Gly Pro Ala Ala Cys Val Gly Arg Ser
 225 230 235 240

Ala Ala Ser Thr Ala Ser Ser Arg Trp Arg Val Ala Gly Ala Thr Xaa
 245 250 255

50 Gly Pro Val Thr Ala Ser Gly Ser Gln Ser Thr Ser Ala Ser Ser Trp
 260 265 270

Trp Asp Leu Gly Cys Met Asp Pro Ser Thr Gly Pro Pro Thr Thr Lys
 55 275 280 285

(2) INFORMATION FOR SEQ ID NO: 138:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 114 amino acids

5 (B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

10	Met Pro Arg Cys Arg Trp Leu Ser Leu Ile Leu Leu Thr Ile Pro Leu	1	5	10	15
	Ala Leu Val Ala Arg Lys Asp Pro Lys Lys Asn Glu Thr Gly Val Leu	20	25	30	
15	Arg Lys Leu Lys Pro Val Asn Ala Phe Xaa Cys Gln Arg Gly Ser Ser	35	40	45	
	Val Xaa Gly Phe Ala Met Gln Glu Tyr Asn Lys Glu Ser Glu Asp Lys	50	55	60	
20	Tyr Val Phe Leu Val Val Lys Thr Leu Gln Ala Gln Leu Gln Val Thr	65	70	75	80
	Asn Leu Leu Glu Tyr Leu Ile Asp Val Glu Ile Ala Arg Ser Asp Cys	85	90	95	
25	Arg Lys Pro Leu Ser Thr Asn Glu Ile Ala Pro Phe Lys Xaa Thr Pro	100	105	110	
30	Ser Xaa				

35 (2) INFORMATION FOR SEQ ID NO: 139:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 120 amino acids

40 (B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

45	Met Ser Pro His Pro Thr Ala Leu Leu Gly Leu Val Leu Cys Leu Ala	1	5	10	15
	Gln Thr Ile His Thr Gln Glu Glu Asp Leu Pro Arg Pro Ser Ile Ser	20	25	30	
50	Ala Glu Pro Gly Thr Val Ile Pro Leu Gly Ser His Val Thr Phe Val	35	40	45	
	Cys Arg Gly Pro Val Gly Val Gln Thr Phe Arg Leu Glu Arg Glu Ser	50	55	60	
55	Arg Ser Thr Tyr Asn Asp Thr Glu Asp Val Ser Gln Ala Ser Pro Ser	65	70	75	80
	Glu Ser Glu Ala Arg Phe Arg Ile Asp Ser Val Ser Glu Gly Asn Ala	85	90	95	
60					

Gly Pro Tyr Arg Cys Ile Tyr Tyr Lys Pro Pro Lys Trp Ser Glu Gln
 100 105 110

5 Ser Asp Tyr Trp Ser Cys Trp Xaa
 115 120

10 (2) INFORMATION FOR SEQ ID NO: 140:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 438 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

Met Asn Thr Pro Asn Gly Asn Ser Leu Ser Ala Ala Glu Leu Thr Cys
 1 5 10 15

20 Gly Met Ile Met Cys Leu Ala Arg Gln Ile Pro Gln Ala Thr Ala Ser
 20 25 30

Met Lys Asp Gly Lys Trp Glu Arg Lys Lys Phe Met Gly Thr Glu Leu
 35 40 45

25 Asn Gly Lys Thr Leu Gly Ile Leu Gly Leu Gly Arg Ile Gly Arg Glu
 50 55 60

30 Val Ala Thr Arg Met Gln Ser Phe Gly Met Lys Thr Ile Gly Tyr Asp
 65 70 75 80

Pro Ile Ile Ser Pro Glu Val Ser Ala Ser Phe Gly Val Gln Gln Leu
 85 90 95

35 Pro Leu Glu Glu Ile Trp Pro Leu Cys Asp Phe Ile Thr Val His Thr
 100 105 110

40 Pro Leu Leu Pro Ser Thr Thr Gly Leu Leu Asn Asp Asn Thr Phe Ala
 115 120 125

45 Gln Cys Lys Lys Gly Val Arg Val Val Asn Cys Ala Arg Gly Gly Ile
 130 135 140

50 Val Asp Glu Gly Ala Leu Leu Arg Ala Leu Gln Ser Gly Gln Cys Ala
 145 150 155 160

55 Gly Ala Ala Leu Asp Val Phe Thr Glu Glu Pro Pro Arg Asp Arg Ala
 165 170 175

60 Leu Val Asp His Glu Asn Val Ile Ser Cys Pro His Leu Gly Ala Ser
 180 185 190

Thr Lys Glu Ala Gln Ser Arg Cys Gly Glu Glu Ile Ala Val Gln Phe
 195 200 205

Val Asp Met Val Lys Gly Lys Ser Leu Thr Gly Val Val Asn Ala Gln
 210 215 220

65 Ala Leu Thr Ser Ala Phe Ser Pro His Thr Lys Pro Trp Ile Gly Leu
 225 230 235 240

Ala Glu Ala Leu Gly Thr Leu Met Arg Ala Trp Ala Gly Ser Pro Lys
 245 250 255

5 Gly Thr Ile Gln Val Ile Thr Gln Gly Thr Ser Leu Lys Asn Ala Gly
 260 265 270

Asn Cys Leu Ser Pro Ala Val Ile Val Gly Leu Leu Lys Glu Ala Ser
 275 280 285

10 Lys Gln Ala Asp Val Asn Leu Val Asn Ala Lys Leu Leu Val Lys Glu
 290 295 300

15 Ala Gly Leu Asn Val Thr Thr Ser His Ser Pro Ala Ala Pro Gly Glu
 305 310 315 320

Gln Gly Phe Gly Glu Cys Leu Leu Ala Val Ala Leu Ala Gly Ala Pro
 325 330 335

20 Tyr Gln Ala Val Gly Leu Val Gln Gly Thr Thr Pro Val Leu Gln Gly
 340 345 350

Leu Asn Gly Ala Val Phe Arg Pro Glu Val Pro Leu Arg Arg Asp Leu
 355 360 365

25 Pro Leu Leu Leu Phe Arg Thr Gln Thr Ser Asp Pro Ala Met Leu Pro
 370 375 380

Thr Met Ile Gly Leu Leu Ala Glu Ala Gly Val Arg Leu Leu Ser Tyr
 385 390 395 400

Gln Thr Ser Leu Val Ser Asp Gly Glu Thr Trp His Val Met Gly Ile
 405 410 415

35 Ser Ser Leu Leu Pro Ser Leu Glu Ala Trp Lys Gln His Val Thr Glu
 420 425 430

Ala Phe Gln Phe His Phe
 435

40

(2) INFORMATION FOR SEQ ID NO: 141:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 164 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

50 Met Ser Arg Pro Thr His Thr Pro Leu Ser Pro Ala Thr Ile Ser Pro
 1 5 10 15

Thr Ile Thr Val Ala Val Phe Phe Ala Val Phe Val Ala Ala Ala
 55 20 25 30

Ala Thr Ala Val Val Ala Ala Ala Thr Thr Ser Ser Gly Arg
 35 40 45

60 Arg Thr Xaa Asp Lys Ser Pro Ile Ala Thr Gln Ser Ser Val Thr His

	50	55	60	
	Ile Ala Ala Lys Arg Cys His Asn Tyr Thr Glu Cys Leu Ser Leu Ile			
	65	70	75	80
5	Arg Xaa Thr Arg Ile Pro Thr Trp Xaa Xaa Xaa Thr Thr Cys Pro Ser			
	85	90	95	
	Arg Ile Pro Ser Thr His Val Ala Ala Gly Ala Gly Phe Ile Arg Glu			
10	100	105	110	
	Arg Ala Cys Leu Gln Cys Gly Ala Val Gly Pro Pro Gly Cys Ile Leu			
	115	120	125	
15	Ala Ser Leu Pro Pro Pro Ser Leu Tyr Leu Ser Pro Glu Leu Arg Cys			
	130	135	140	
	Met Pro Lys Arg Val Glu Ala Arg Ser Glu Leu Arg Leu Cys Pro Pro			
	145	150	155	160
20	Gly Val Xaa Xaa			

25

(2) INFORMATION FOR SEQ ID NO: 142:

	(i) SEQUENCE CHARACTERISTICS:			
	(A) LENGTH: 73 amino acids			
30	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 142:			
	Met Gln Arg Trp Val Cys Ile Leu Glu Phe Lys Glu Asn Leu Phe Gln			
	1	5	10	15
	Ile Pro Ser Ser Leu Val Ala Leu Leu Asn Thr Leu Phe Leu Asp Ile			
	20	25	30	
40	Leu His Pro Gln Asn Ser Leu Ser Pro His Gly Ser Phe Ser Leu Ser			
	35	40	45	
	Ser Leu Ser Phe Pro Pro Leu Pro Val Ser Ser Leu Gln Pro Phe Leu			
	50	55	60	
45	Phe Leu Arg Ser Leu Leu Cys Arg Xaa			
	65	70		

50

(2) INFORMATION FOR SEQ ID NO: 143:

	(i) SEQUENCE CHARACTERISTICS:			
	(A) LENGTH: 123 amino acids			
55	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 143:			
	Phe Gly Thr Arg Phe Leu Ala Asn Leu Leu Glu Glu Asp Asn Lys			
	60	1	5	10
				15

Phe Cys Ala Asp Cys Gln Ser Lys Gly Pro Arg Trp Ala Ser Trp Asn
 20 25 30

5 Ile Gly Val Phe Ile Cys Ile Arg Cys Ala Xaa Ile His Arg Asn Leu
 35 40 45

Gly Val His Ile Ser Arg Val Lys Ser Val Asn Leu Asp Gln Trp Thr
 50 55 60

10 Gln Val Gln Ile Gln Cys Met Gln Xaa Met Gly Asn Gly Lys Ala Asn
 65 70 75 80

15 Arg Leu Tyr Glu Ala Tyr Leu Pro Glu Thr Phe Arg Arg Pro Gln Ile
 85 90 95

Asp Pro Ala Val Glu Gly Phe Ile Arg Asp Xaa Tyr Glu Lys Lys
 100 105 110

20 Tyr Met Asp Arg Ser Leu Gly His Gln Cys Leu
 115 120

25 (2) INFORMATION FOR SEQ ID NO: 144:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 138 amino acids

(B) TYPE: amino acid

(C) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

Met Ser Leu Tyr Asp Asp Leu Gly Val Glu Thr Ser Asp Ser Lys Thr
 1 5 10 15

35 Glu Gly Trp Ser Lys Asn Phe Lys Leu Leu Gln Ser Gln Leu Gln Val
 20 25 30

40 Lys Lys Ala Ala Leu Thr Gln Ala Lys Ser Gln Arg Thr Lys Gln Ser
 35 40 45

Thr Val Leu Ala Pro Val Ile Asp Leu Lys Arg Gly Gly Ser Ser Asp
 50 55 60

45 Asp Arg Gln Ile Val Asp Thr Pro Pro His Val Ala Ala Gly Leu Lys
 65 70 75 80

50 Asp Pro Val Pro Ser Gly Phe Ser Ala Gly Glu Val Leu Ile Pro Leu
 85 90 95

Ala Asp Glu Tyr Asp Pro Met Phe Pro Asn Asp Tyr Glu Lys Val Val
 100 105 110

55 Lys Arg Ala Lys Arg Gly Thr Thr Glu Thr Ala Gly Val Xaa Lys Thr
 115 120 125

Lys Gly Asn Arg Arg Lys Gly Lys Lys Ala
 130 135

(2) INFORMATION FOR SEQ ID NO: 145:

5 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 356 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 145:

10 Met Leu Ala Arg Ala Ala Arg Gly Thr Gly Ala Leu Leu Leu Arg Gly
 1 5 10 15

Ser Leu Leu Ala Ser Gly Arg Ala Pro Arg Arg Ala Ser Ser Gly Leu
 20 25 30

15 Pro Arg Asn Thr Val Val Leu Phe Val Pro Gln Gln Glu Ala Trp Val
 35 40 45

20 Val Glu Arg Met Gly Arg Phe His Arg Ile Leu Glu Pro Gly Leu Asn
 50 55 60

Ile Leu Ile Pro Val Leu Asp Arg Ile Arg Tyr Val Gln Ser Leu Lys
 65 70 75 80

25 Glu Ile Val Ile Asn Val Pro Glu Gln Ser Ala Val Thr Leu Asp Asn
 85 90 95

Val Thr Leu Gln Ile Asp Gly Val Leu Tyr Leu Arg Ile Met Asp Pro
 100 105 110

30 Tyr Lys Ala Ser Tyr Gly Val Glu Asp Pro Glu Tyr Ala Val Thr Gln
 115 120 125

Leu Ala Gln Thr Thr Met Arg Ser Glu Leu Gly Lys Leu Ser Leu Asp
 35 130 135 140

Lys Val Phe Arg Glu Arg Glu Ser Leu Asn Ala Ser Ile Val Asp Ala
 145 150 155 160

40 Ile Asn Gln Ala Ala Asp Cys Trp Gly Ile Arg Cys Leu Arg Tyr Glu
 165 170 175

Ile Lys Asp Ile His Val Pro Pro Arg Val Lys Glu Ser Met Gln Met
 180 185 190

45 Gln Val Glu Ala Glu Arg Arg Lys Arg Ala Thr Val Leu Glu Ser Glu
 195 200 205

Gly Thr Arg Glu Ser Ala Ile Asn Val Ala Glu Gly Lys Lys Gln Ala
 50 210 215 220

Gln Ile Leu Ala Ser Glu Ala Glu Lys Ala Glu Gln Ile Asn Gln Ala
 225 230 235 240

55 Ala Gly Glu Ala Ser Ala Val Leu Ala Lys Ala Lys Ala Lys Ala Glu
 245 250 255

Ala Ile Arg Ile Leu Ala Ala Leu Thr Gln His Asn Gly Asp Ala
 260 265 270

60

Ala Ala Ser Leu Thr Val Ala Glu Gln Tyr Val Ser Ala Phe Ser Lys
275 280 285

5 Leu Ala Lys Asp Ser Asn Thr Ile Leu Leu Pro Ser Asn Pro Gly Asp
290 295 300

Val Thr Ser Met Val Ala Gln Ala Met Gly Val Tyr Gly Ala Leu Thr
 305 310 315 320

10 Lys Ala Pro Val Pro Gly Thr Pro Asp Ser Leu Ser Ser Gly Ser Ser
325 330 335

Arg Asp Val Gln Gly Thr Asp Ala Ser Leu Asp Glu Glu Leu Asp Arg
340 345 350

15 Val Lys Met Ser
355

20

(2) INFORMATION FOR SEQ ID NO: 146:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

Met Tyr Ile Leu Leu Phe Trp Gly Gly Xaa Phe His Arg Cys Leu Ser
30 1 5 10 15 20 25 30 35 40 45 50

Xaa Leu Phe Asp Pro Glu Leu Xaa Ser Xaa Pro Gly Ile Ser Xaa Phe
 20 25 30

35 Thr Val Xaa Leu Gln Met Thr Xaa
35 40

40 (2) INFORMATION FOR SEQ ID NO: 147:

(i) SEQUENCE CHARACTERISTICS:-

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

Met Pro Ser Pro Lys Tyr Cys Met His Thr Asn Asp Val Gln Ser Val
1 5 10 15

Glu Tyr Asn Gly Asp Thr Leu Phe Gln Lys Leu Ser Ser Ser Xaa Leu
20 25 30

55 Ser Phe Lys Ser Ile His Ile Tyr Pro Asn Glu Xaa Lys Thr Cys Xaa
 35 40 45

Xaa Ile Phe Ile Ser Lys Val Tyr Met Ile Ser Lys Thr Trp Lys Xaa
50 55 60

60 Pro Arg Phe Thr Ser Xaa Gly

65 70

5 (2) INFORMATION FOR SEQ ID NO: 148:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

Met Asn Phe Val Leu Phe Phe Ile Gly Ile Asn Val Gly Cys Arg Gly
1 5 10 15

15 Glu Asn Ser Leu Lys Tyr Phe Thr Val Thr Val Leu Cys Ser Pro Arg
20 25 30

20 Asp

25 (2) INFORMATION FOR SEQ ID NO: 149:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 78 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

Met Lys Glu Ala Gly Lys Gly Gly Val Ala Asp Ser Arg Glu Leu Lys
1 5 10 15

35 Pro Met Val Gly Gly Asp Glu Glu Val Ala Ala Leu Gln Glu Phe His
20 25 30

Phe His Phe Leu Ser Leu Ser Val Phe Thr Asp Cys Thr Ser Ser Gly
35 40 45

40 Glu Ala Phe Val Ile Cys Ile Thr Gln Thr Cys Cys Ser Phe Cys Leu
50 55 60

45 Cys Ala Tyr Pro Ser Leu Gly Trp Gln Asn Ser Cys His Asn
65 70 75

50 (2) INFORMATION FOR SEQ ID NO: 150:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

Met Phe Ser Ser Lys Ser Leu Leu Val Pro Phe Cys Phe Arg Ser
1 5 10 15

60 Ala Ala His Leu Glu Leu Ser Val Trp Cys Val Cys Gly Val Arg Xaa

20

25

30

5

(2) INFORMATION FOR SEQ ID NO: 151:

10 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 464 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

15

Met Leu Ala Leu Gly Asn Asn His Phe Ile Gly Phe Val Asn Asp Ser
1 5 10 15

20

Val Thr Lys Ser Ile Val Ala Leu Arg Leu Thr Leu Val Val Lys Val
20 25 30Ser Thr Xaa Pro Gly Glu Ser His Ala Asn Asp Leu Glu Cys Ser Gly
35 40 45

25

Lys Gly Lys Cys Thr Thr Lys Pro Ser Glu Ala Thr Phe Ser Cys Thr
50 55 60Cys Glu Glu Gln Tyr Val Gly Thr Phe Cys Glu Glu Tyr Asp Ala Cys
65 70 75 80

30

Gln Arg Lys Pro Cys Gln Asn Asn Ala Ser Cys Ile Asp Ala Asn Glu
85 90 95

35

Lys Gln Asp Gly Ser Asn Phe Thr Cys Val Cys Leu Pro Gly Tyr Thr
100 105 110Gly Glu Leu Cys Gln Ser Lys Ile Asp Tyr Cys Ile Leu Asp Pro Cys
115 120 125

40

Arg Asn Gly Ala Thr Cys Ile Ser Ser Leu Ser Gly Phe Thr Cys Gln
130 135 140Cys Pro Glu Gly Tyr Phe Gly Ser Ala Cys Glu Glu Lys Val Asp Pro
145 150 155 160

45

Cys Ala Ser Ser Pro Cys Gln Asn Asn Gly Thr Cys Tyr Val Asp Gly
165 170 175

50

Val His Phe Thr Cys Asn Cys Ser Pro Gly Phe Thr Gly Pro Thr Cys
180 185 190Ala Gln Leu Ile Asp Phe Cys Ala Leu Ser Pro Cys Ala His Gly Thr
195 200 205

55

Cys Arg Ser Val Gly Thr Ser Tyr Lys Cys Leu Cys Asp Pro Gly Tyr
210 215 220His Gly Leu Tyr Cys Glu Glu Glu Tyr Asn Glu Cys Leu Ser Ala Pro
225 230 235 240

60

Cys Leu Asn Ala Ala Thr Cys Arg Asp Leu Val Asn Gly Tyr Glu Cys
 245 250 255

5 Val Cys Leu Ala Glu Tyr Lys Gly Thr His Cys Glu Leu Tyr Lys Asp
 260 265 270

Pro Cys Ala Asn Val Ser Cys Leu Asn Gly Ala Thr Cys Asp Ser Asp
 275 280 285

10 Gly Leu Asn Gly Thr Cys Ile Cys Ala Pro Gly Phe Thr Gly Glu Glu
 290 295 300

Cys Asp Ile Asp Ile Asn Glu Cys Asp Ser Asn Pro Cys His His Gly
 305 310 315 320

15 Gly Ser Cys Leu Asp Gln Pro Asn Gly Tyr Asn Xaa His Cys Pro His
 325 330 335

Gly Trp Val Gly Ala Asn Cys Glu Ile His Leu Gln Trp Lys Ser Gly
 20 340 345 350

His Met Ala Glu Ser Leu Thr Asn Met Pro Arg His Ser Leu Tyr Ile
 355 360 365

25 Ile Ile Gly Ala Leu Cys Val Ala Phe Ile Leu Met Leu Ile Ile Leu
 370 375 380

Ile Val Gly Ile Cys Arg Ile Ser Arg Ile Glu Tyr Gln Gly Ser Ser
 385 390 395 400

30 Arg Pro Ala Tyr Xaa Glu Phe Tyr Asn Cys Arg Ser Ile Asp Ser Glu
 405 410 415

Phe Ser Asn Ala Ile Ala Ser Ile Arg His Ala Arg Phe Gly Lys Lys
 35 420 425 430

Ser Arg Pro Ala Met Tyr Asp Val Ser Pro Ile Ala Tyr Glu Asp Tyr
 435 440 445

40 Ser Pro Asp Asp Lys Pro Leu Val Thr Leu Ile Lys Thr Lys Asp Leu
 450 455 460

45

(2) INFORMATION FOR SEQ ID NO: 152:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 151 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

55 Met His His Gln Met Thr Arg Thr Thr Leu Met Thr Lys Gln His Glu
 1 5 10 15

60 Leu Gly Gly Leu Leu Ala Leu Val Gln Asn Cys Gln Ser Glu Met Asn
 20 25 30

	Ile Lys Asp Ser Arg Ala Val Gly Leu Ser Val Lys Arg Leu Cys Ile
	35 40 45
5	Ser Phe Val Asp Glu Phe Cys Glu Arg Thr Glu Arg Pro Leu Tyr Leu
	50 55 60
	Ala Gln Gly Leu Phe Met Lys Arg Glu Thr Tyr Trp Glu Val Gln Asp
	65 70 75 80
10	Ser Gly Ile Ser Pro Leu Leu Leu Leu Ser Thr Ala Leu Asp Cys
	85 90 95
15	Ser Pro Glu Ala Glu Thr Arg Gln Ser Pro Gly Gly Arg Lys Met Leu
	100 105 110
	Gln Glu Pro Thr Leu Ser Met Ser Leu Gln Ile Leu Thr Gly Phe Leu
	115 120 125
20	Trp Val Gln Leu Trp Asn Trp Glu Thr Phe Leu Arg Ile Arg Thr His
	130 135 140
	Ser Thr Asp Ala Ser Cys Pro
	145 150
25	

(2) INFORMATION FOR SEQ ID NO: 153:

30	(i) SEQUENCE CHARACTERISTICS:
	(A) LENGTH: 299 amino acids
	(B) TYPE: amino acid
	(D) TOPOLOGY: linear
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 153:
	Met Ala Gln Asn Leu Lys Asp Leu Ala Gly Arg Leu Pro Ala Gly Pro
	1 5 10 15
40	Arg Gly Met Gly Thr Ala Leu Lys Leu Leu Leu Gly Ala Gly Ala Val
	20 25 30
	Ala Tyr Gly Val Arg Glu Ser Val Phe Thr Val Glu Gly Gly His Arg
	35 40 45
45	Ala Ile Phe Phe Asn Arg Ile Gly Gly Val Gln Gln Asp Thr Ile Leu
	50 55 60
	Ala Glu Gly Leu His Phe Arg Ile Pro Trp Phe Gln Tyr Pro Ile Ile
	65 70 75 80
50	Tyr Asp Ile Arg Ala Arg Pro Arg Lys Ile Ser Ser Pro Thr Gly Ser
	85 90 95
	Lys Asp Leu Gln Met Val Asn Ile Ser Leu Arg Val Leu Ser Arg Pro
55	100 105 110
	Asn Ala Gln Glu Leu Pro Ser Met Tyr Gln Arg Leu Gly Leu Asp Tyr
	115 120 125
60	Glu Glu Arg Val Leu Pro Ser Ile Val Asn Glu Val Leu Lys Ser Val

	130	135	140	
	Val Ala Lys Phe Asn Ala Ser Gln Leu Ile Thr Gln Arg Ala Gln Val			
	145	150	155	160
5	Ser Leu Leu Ile Arg Arg Glu Leu Thr Glu Arg Ala Lys Asp Phe Ser			
	165	170	175	
10	Leu Ile Leu Asp Asp Val Ala Ile Thr Glu Leu Ser Phe Ser Arg Glu			
	180	185	190	
	Tyr Thr Ala Ala Val Glu Ala Lys Gln Val Ala Gln Gln Glu Ala Gln			
	195	200	205	
15	Arg Ala Xaa Phe Leu Val Glu Lys Ala Lys Gln Glu Gln Arg Gln Lys			
	210	215	220	
	Ile Val Gln Ala Glu Gly Glu Ala Glu Ala Ala Lys Met Leu Gly Glu			
	225	230	235	240
20	Ala Leu Ser Lys Asn Pro Gly Tyr Ile Lys Leu Arg Lys Ile Arg Ala			
	245	250	255	
25	Ala Gln Asn Ile Ser Lys Thr Ile Ala Thr Ser Gln Asn Arg Ile Tyr			
	260	265	270	
	Leu Thr Ala Asp Asn Leu Val Leu Asn Leu Gln Asp Glu Ser Phe Thr			
	275	280	285	
30	Arg Gly Ser Asp Ser Leu Ile Lys Gly Lys Lys			
	290	295		

35 (2) INFORMATION FOR SEQ ID NO: 154:

	(i) SEQUENCE CHARACTERISTICS:			
	(A) LENGTH: 398 amino acids			
	(B) TYPE: amino acid			
40	(D) TOPOLOGY: linear			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 154:			
	Met Leu Arg Gly Pro Trp Arg Gln Leu Trp Leu Phe Xaa Leu Leu Leu			
	1	5	10	15
45	Leu Pro Gly Ala Pro Glu Pro Arg Gly Ala Ser Arg Pro Trp Glu Gly			
	20	25	30	
50	Thr Asp Glu Pro Gly Ser Ala Trp Ala Trp Pro Gly Phe Gln Arg Leu			
	35	40	45	
	Gln Glu Gln Leu Arg Ala Ala Gly Ala Leu Ser Lys Arg Tyr Trp Thr			
	50	55	60	
55	Leu Phe Ser Cys Gln Val Trp Pro Asp Asp Cys Asp Glu Asp Glu Glu			
	65	70	75	80
60	Ala Ala Thr Gly Pro Leu Gly Trp Arg Leu Pro Leu Leu Gly Gln Arg			
	85	90	95	

Tyr Leu Asp Leu Leu Thr Thr Trp Tyr Cys Ser Phe Lys Asp Cys Cys
 100 105 110

5 Pro Arg Gly Asp Cys Arg Ile Ser Asn Asn Phe Thr Gly Leu Glu Trp
 115 120 125

Asp Leu Asn Val Arg Leu His Gly Gln His Leu Val Gln Gln Leu Val
 130 135 140

10 Leu Arg Thr Val Arg Gly Tyr Leu Glu Thr Pro Gln Pro Glu Lys Ala
 145 150 155 160

Leu Ala Leu Ser Phe His Gly Trp Ser Gly Thr Gly Lys Asn Phe Val
 165 170 175

15 Ala Arg Met Leu Val Glu Asn Leu Tyr Arg Asp Gly Leu Met Ser Asp
 180 185 190

20 Cys Val Arg Met Phe Ile Ala Thr Phe His Phe Pro His Pro Lys Tyr
 195 200 205

Val Asp Leu Tyr Lys Glu Gln Leu Met Ser Gln Ile Arg Glu Thr Gln
 210 215 220

25 Gln Leu Cys His Gln Thr Leu Phe Ile Phe Asp Glu Ala Glu Lys Leu
 225 230 235 240

His Pro Gly Leu Leu Glu Val Leu Gly Pro His Leu Glu Arg Arg Ala
 245 250 255

30 Pro Xaa Gly His Arg Ala Glu Ser Pro Trp Thr Ile Phe Leu Phe Leu
 260 265 270

35 Ser Asn Leu Arg Gly Asp Ile Ile Asn Glu Val Val Leu Lys Leu Leu
 275 280 285

Lys Ala Gly Trp Ser Arg Glu Glu Ile Thr Met Glu His Leu Glu Pro
 290 295 300

40 His Leu Gln Ala Glu Ile Val Glu Thr Ile Asp Asn Gly Phe Gly His
 305 310 315 320

Ser Arg Leu Val Lys Glu Asn Leu Ile Asp Tyr Phe Ile Pro Phe Leu
 325 330 335

45 Pro Leu Glu Tyr Arg His Val Arg Leu Cys Ala Arg Asp Ala Phe Leu
 340 345 350

50 Ser Gln Glu Leu Leu Tyr Lys Glu Glu Thr Leu Asp Glu Ile Ala Gln
 355 360 365

Met Met Val Tyr Val Pro Lys Glu Glu Gln Leu Phe Ser Ser Gln Gly
 370 375 380

55 Cys Lys Ser Ile Ser Gln Arg Ile Asn Tyr Phe Leu Ser Xaa
 385 390 395

60 (2) INFORMATION FOR SEQ ID NO: 155:
 --

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 83 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

Met Ala Phe Thr Leu Tyr Ser Leu Leu Gln Ala Xaa Leu Leu Cys Val
1 5 10 15

10 Asn Ala Ile Ala Val Leu His Glu Glu Arg Phe Leu Lys Asn Ile Gly
20 25 30

15 Trp Gly Thr Asp Gln Gly Ile Gly Gly Phe Gly Glu Glu Pro Gly Ile
35 40 45

Lys Ser Gln Leu Met Asn Leu Ile Arg Ser Val Arg Thr Val Met Arg
50 55 60

20 Val Pro Leu Ile Ile Val Asn Ser Ile Ala Ile Val Leu Leu Leu Leu
65 70 75 80

Phe Gly Xaa

25

(2) INFORMATION FOR SEQ ID NO: 156:

30 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

Met Ala Pro Arg Asn Gln Gly Ser Phe Ser Phe Gly Asn Phe Met Leu
1 5 10 15

40 Phe Leu Val Leu Ile Glu Arg Arg Tyr Leu Pro Phe Leu Ser Pro Ile
20 25 30

Leu Phe Cys Cys Ser Thr His Asn Arg Ser Ala Val Thr Ala Thr Asn
35 40 45

45 Leu Xaa
50

50 (2) INFORMATION FOR SEQ ID NO: 157:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

Met Asp Val Leu Thr Val Ala Phe Leu Ser Ile Leu Ile Thr Ala Pro
1 5 10 15

60

Ile Gly Ser Leu Leu Ile Gly Leu Leu Gly Pro Arg Leu Leu Gln Lys
20 25 30

5 Val Glu His Gln Asn Lys Asp Glu Glu Val Gln Gly Glu Thr Ser Val
35 40 45

Gln Val Xaa
50

10

(2) INFORMATION FOR SEQ ID NO: 158:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

20 Pro Asn Ser Phe Ser Cys Leu Gly Leu Ala Gly Thr Gly Ala Gly Ile
1 5 10 15

Xaa

25

(2) INFORMATION FOR SEQ ID NO: 159:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 53 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

35 Met Gly Arg Tyr His Phe Val Phe Leu Thr Phe Phe Phe Ser Thr Tyr
1 5 10 15

40 Ser Ser Cys Phe Tyr Pro Val Val Ser Gln Val Leu Tyr Leu Val Cys
20 25 30

Ser Cys Thr Ala Asp Arg Pro Leu Met Ala Pro Val Gly Ser Cys Leu
35 40 45

45 Gly Gly Arg Asn Xaa
50

50 (2) INFORMATION FOR SEQ ID NO: 160:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 64 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

55 Met Phe Val Thr Leu Ser Ile Leu Asn Ile Thr Ile Glu Lys Asp Lys
1 5 10 15

Ser Thr Asn Arg Phe Arg Asp Val Phe Leu Gln His Ile Leu Val Ile
 20 25 30

5 Leu Met Pro Ser Leu Thr Tyr Cys Leu Ile Gly Gln His Leu Cys Ser
 35 40 45

Phe Thr Arg Tyr Val Ser Leu Cys Tyr Ser Arg Cys His Ser Trp Xaa
 50 55 60

10

15 (2) INFORMATION FOR SEQ ID NO: 161:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

20 Met Ser Ile Cys Pro Leu Leu Val Met Leu Ile Leu Ile Thr Trp Val
 1 5 10 15

25 Arg Cys Pro Val Ser Pro Val Tyr Arg Tyr Cys Phe Ser Phe Cys Asn
 20 25 30

30 Xaa

35 (2) INFORMATION FOR SEQ ID NO: 162:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 95 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

Met Gln Asp Ile Val Tyr Lys Leu Val Pro Gly Leu Gln Glu Gly Glu
 1 5 10 15

45 Cys Leu Thr Val Leu Leu Ile Pro Glu Val Pro Ala Trp Pro Leu Gln
 20 25 30

50 Pro Leu Leu Ser Trp Lys Phe Gly Ser Arg Met Gly Gly Pro Phe Pro
 35 40 45

55 Phe Gly Arg Ile Thr Val Phe Ser Ser Leu Leu Ser Ala Gln Leu His
 50 55 60

55 Leu Leu Gly Trp Ser Leu Leu Ser Ser Lys Met Arg Xaa His Leu Phe
 65 70 75 80

60 Thr Pro Tyr Val Tyr Ser Phe Ser Lys Tyr Gly Ser His Val Xaa
 85 90 95

60

(2) INFORMATION FOR SEQ ID NO: 163:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

10 Met Lys Val Leu Ala Thr Ser Phe Val Leu Gly Ser Leu Gly Leu Ala
1 5 10 15

Phe Tyr Leu Pro Leu Val Val Thr Thr Pro Lys Thr Leu Ala Ile Pro
20 25 30

15 Xaa Glu Ala Ala Arg Ser Cys Gly Glu Ser Tyr His Gln Cys His Asn
35 40 45

Leu Tyr Cys His Leu Trp Pro Trp Leu Xaa
20 50 55

(2) INFORMATION FOR SEQ ID NO: 164:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

Met Asp Tyr Gly Tyr Tyr Ser Ala Gly Gln Phe Leu Leu His Leu Phe
1 5 10 15

35 Leu Ala Asp Leu Thr Gln Ala Thr Thr Gln Gln Lys Thr Asn Thr Ser
20 25 30

Glu Asn Gly Cys Lys Phe Val Cys Ala Val Phe Xaa
35 40

40

(2) INFORMATION FOR SEQ ID NO: 165:

45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

50 Gly Ile Val Leu Leu Ile Gly Val Leu Val Gln Val Ser Ala Val Asp
1 5 10 15

55 Asp Xaa

60 (2) INFORMATION FOR SEQ ID NO: 166:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

Met Gly Asn Ala Phe Glu Val Thr Gly Leu Met Leu Ala Leu Leu Cys
1 5 10 15

10 Tyr Val Val Asp Gly Gln Lys Pro Lys Xaa Gly Phe Xaa Xaa
20 25 30

15 (2) INFORMATION FOR SEQ ID NO: 167:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

Met Ser His Glu Lys Ser Asn Glu Leu Val Leu Leu Ile Val Thr Val
1 5 10 15

25 Met Arg Ser Leu Thr Tyr Asn Ile Ala Val Val Ala Ala Trp Phe Asn
20 25 30

30 Gly Cys Ile Arg Xaa
35

35 (2) INFORMATION FOR SEQ ID NO: 168:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

Met Tyr Leu Leu Tyr Leu Pro Ser Ala Leu Leu Pro Pro Tyr Pro Thr
1 5 10 15

45 Cys Pro Tyr Glu His Gly Ser Pro Trp Pro His Thr Pro Ala Lys Leu
20 25 30

Leu Cys Cys Phe Ala Phe Leu Xaa
35 40

50 (2) INFORMATION FOR SEQ ID NO: 169:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 169:
60

Met Lys Phe Ile Val Trp Arg Arg Phe Lys Trp Val Ile Ile Gly Leu
1 5 10 15

5 Leu Phe Leu Leu Ile Leu Leu Phe Val Ala Val Leu Leu Tyr Ser
20 25 30

Leu Pro Asn Tyr Leu Ser Met Lys Ile Val Lys Pro Asn Val Xaa
35 40 45

10

(2) INFORMATION FOR SEQ ID NO: 170:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

20 Ile Glu Trp Ser Gly Tyr Asn Lys Pro Glu Arg Lys Gly Pro Leu Ala
1 5 10 15

25 Leu Phe Leu Val Phe Leu Asp Thr Pro Pro Leu Gln Gly Asp
20 25 30

Leu Xaa

30

(2) INFORMATION FOR SEQ ID NO: 171:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

40 Met Ser Leu Leu Xaa
1 5

45 (2) INFORMATION FOR SEQ ID NO: 172:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

Met Gln Leu Leu Ile Val Trp Asn Glu Ser Leu Thr Asn Ser Val Pro
1 5 10 15

55 Ala Ser Val Asp Thr Ser Gln Cys Xaa
20 25

60 (2) INFORMATION FOR SEQ ID NO: 173:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 262 amino acids
- (B) TYPE: amino acid
- 5 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

	Met Ala Leu Gly Leu Lys Cys Phe Arg Met Val His Pro Thr Phe Arg		
1	5	10	15
10	Asn Tyr Leu Ala Ala Ser Ile Arg Pro Val Ser Glu Val Thr Leu Lys		
	20	25	30
15	Thr Val His Glu Arg Gln His Gly His Arg Gln Tyr Met Ala Tyr Ser		
	35	40	45
20	Ala Val Pro Val Arg His Phe Ala Thr Lys Lys Ala Lys Ala Lys Gly		
	50	55	60
25	Lys Gly Gln Ser Gln Thr Arg Val Asn Ile Asn Ala Ala Leu Val Glu		
	65	70	75
	Asp Ile Ile Asn Leu Glu Glu Val Asn Glu Glu Met Lys Ser Val Ile		
	85	90	95
30	Glu Ala Leu Lys Asp Asn Phe Asn Lys Thr Leu Asn Ile Arg Thr Ser		
	100	105	110
35	Pro Gly Ser Leu Asp Lys Ile Ala Val Val Thr Ala Asp Gly Lys Leu		
	115	120	125
40	Ala Leu Asn Gln Ile Ser Gln Ile Ser Met Lys Ser Pro Gln Leu Ile		
	130	135	140
45	Leu Val Asn Met Ala Ser Phe Pro Glu Cys Thr Ala Ala Ala Ile Lys		
	145	150	155
	Ala Ile Arg Glu Ser Gly Met Asn Leu Asn Pro Glu Val Glu Gly Thr		
	165	170	175
50	Leu Ile Arg Val Pro Ile Pro Gln Val Thr Arg Glu His Arg Glu Met		
	180	185	190
55	Leu Val Lys Leu Ala Lys Gln Asn Thr Asn Lys Ala Lys Asp Ser Leu		
	195	200	205
	Arg Lys Val Arg Thr Asn Ser Met Asn Lys Leu Lys Lys Ser Lys Asp		
	210	215	220
60	Thr Val Ser Glu Asp Thr Ile Arg Leu Ile Glu Lys Gln Ile Ser Gln		
	225	230	235
	Met Ala Asp Asp Thr Val Ala Glu Leu Asp Arg His Leu Ala Val Lys		
	245	250	255
	Thr Lys Glu Leu Leu Gly		
	260		

(2) INFORMATION FOR SEQ ID NO: 174:

(i) SEQUENCE CHARACTERISTICS:

5 (A) LENGTH: 967 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

10 Met Gln Arg Ala Val Pro Glu Gly Phe Gly Arg Arg Lys Leu Gly Ser
 1 5 10 15

Asp Met Gly Asn Ala Glu Arg Ala Pro Gly Ser Arg Ser Phe Gly Pro
 20 25 30

15 Val Pro Thr Leu Leu Leu Xaa Ala Ala Leu Leu Xaa Val Ser Asp
 35 40 45

Ala Leu Gly Arg Pro Ser Glu Glu Asp Glu Glu Leu Val Val Pro Glu
 50 55 60

20 Leu Glu Arg Ala Pro Gly His Gly Thr Thr Arg Leu Arg Leu His Ala
 65 70 75 80

25 Phe Asp Gln Gln Leu Asp Leu Glu Leu Arg Pro Asp Ser Ser Phe Leu
 85 90 95

Ala Pro Gly Phe Thr Leu Gln Asn Val Gly Arg Lys Ser Gly Ser Glu
 100 105 110

30 Thr Pro Leu Pro Glu Thr Asp Leu Ala His Cys Phe Tyr Ser Gly Thr
 115 120 125

Val Asn Gly Asp Pro Ser Ser Ala Ala Ala Leu Ser Leu Cys Glu Gly
 130 135 140

35 Val Arg Gly Ala Phe Tyr Leu Leu Gly Glu Ala Tyr Phe Ile Gln Pro
 145 150 155 160

40 Leu Pro Ala Ala Ser Glu Arg Leu Xaa Thr Ala Ala Pro Gly Glu Lys
 165 170 175

Pro Pro Ala Pro Leu Gln Phe His Leu Leu Arg Arg Asn Arg Gln Gly
 180 185 190

45 Asp Val Gly Gly Thr Cys Gly Val Val Asp Asp Glu Pro Arg Pro Thr
 195 200 205

Gly Lys Ala Glu Thr Glu Asp Glu Gly Thr Glu Gly Glu Asp
 210 215 220

50 Glu Gly Pro Gln Trp Ser Pro Gln Asp Pro Ala Leu Gln Gly Val Gly
 225 230 235 240

Gln Pro Thr Gly Thr Gly Ser Ile Arg Lys Lys Arg Phe Val Ser Ser
 245 250 255

His Arg Tyr Val Glu Thr Met Leu Val Ala Asp Gln Ser Met Ala Glu
 260 265 270

60 Phe His Gly Ser Gly Leu Lys His Tyr Leu Leu Thr Leu Phe Ser Val

	275	280	285
	Ala Ala Arg Leu Xaa Lys His Pro Xaa Ile Arg Asn Ser Val Ser Leu		
5	290	295	300
	Val Val Val Lys Ile Leu Val Ile His Asp Glu Gln Lys Gly Pro Glu		
	305	310	315
10	Val Thr Ser Asn Ala Ala Leu Thr Leu Arg Asn Phe Cys Asn Trp Gln		
	325	330	335
	Lys Gln His Asn Pro Pro Ser Asp Arg Asp Ala Glu His Tyr Asp Thr		
	340	345	350
15	Ala Ile Leu Phe Thr Arg Gln Asp Leu Cys Gly Ser Gln Thr Cys Asp		
	355	360	365
	Thr Leu Gly Met Ala Asp Val Gly Thr Val Cys Asp Pro Ser Arg Ser		
20	370	375	380
	Cys Ser Val Ile Glu Asp Asp Gly Leu Gln Ala Ala Phe Thr Thr Ala		
	385	390	395
25	His Glu Leu Gly His Val Phe Asn Met Pro His Asp Asp Ala Lys Gln		
	405	410	415
	Cys Ala Ser Leu Asn Gly Val Asn Gln Asp Ser His Met Met Ala Ser		
	420	425	430
30	Met Leu Ser Asn Leu Asp His Ser Gln Pro Trp Ser Pro Cys Ser Ala		
	435	440	445
	Tyr Met Ile Thr Ser Phe Leu Asp Asn Gly His Gly Glu Cys Leu Met		
35	450	455	460
	Asp Lys Pro Gln Asn Pro Ile Gln Leu Pro Gly Asp Leu Pro Gly Thr		
	465	470	475
40	Ser Tyr Asp Ala Asn Arg Gln Cys Gln Phe Thr Phe Gly Glu Asp Ser		
	485	490	495
	Lys His Cys Pro Asp Ala Ala Ser Thr Cys Ser Thr Leu Trp Cys Thr		
	500	505	510
45	Gly Thr Ser Gly Gly Val Leu Val Cys Gln Thr Lys His Phe Pro Trp		
	515	520	525
	Ala Asp Gly Thr Ser Cys Gly Glu Gly Lys Trp Cys Ile Asn Gly Lys		
50	530	535	540
	Cys Val Xaa Lys Thr Asp Arg Lys His Phe Asp Thr Pro Phe His Gly		
	545	550	555
55	Ser Trp Gly Met Trp Gly Pro Trp Gly Asp Cys Ser Arg Thr Cys Gly		
	565	570	575
	Gly Gly Val Gln Tyr Thr Met Arg Glu Cys Asp Asn Pro Val Pro Lys		
	580	585	590
60	Asn Gly Gly Lys Tyr Cys Glu Gly Lys Arg Val Arg Tyr Arg Ser Cys		

	595	600	605
	Asn Leu Glu Asp Cys Pro Asp Asn Asn Gly Lys Thr Phe Arg Glu Glu		
	610	615	620
5	Gln Cys Glu Ala His Asn Glu Phe Ser Lys Ala Ser Phe Gly Ser Gly		
	625	630	635
	640		
10	Pro Ala Val Glu Trp Ile Pro Lys Tyr Ala Gly Val Ser Pro Lys Asp		
	645	650	655
	Arg Cys Lys Leu Ile Cys Gln Ala Lys Gly Ile Gly Tyr Phe Phe Val		
	660	665	670
15	Leu Gln Pro Lys Val Val Asp Gly Thr Pro Cys Ser Pro Asp Ser Thr		
	675	680	685
	Ser Val Cys Val Gln Gly Gln Cys Val Lys Ala Gly Cys Asp Arg Ile		
	690	695	700
20	Ile Asp Ser Lys Lys Phe Asp Lys Cys Gly Val Cys Gly Gly Asn		
	705	710	715
	720		
25	Gly Ser Thr Cys Lys Ile Ser Gly Ser Val Thr Ser Ala Lys Pro		
	725	730	735
	Gly Tyr His Asp Ile Ile Thr Ile Pro Thr Gly Ala Thr Asn Ile Glu		
	740	745	750
30	Val Lys Gln Arg Asn Gln Arg Gly Ser Arg Asn Asn Gly Ser Phe Leu		
	755	760	765
	Ala Ile Lys Ala Ala Asp Gly Thr Tyr Ile Leu Asn Gly Asp Tyr Thr		
	770	775	780
35	Leu Ser Thr Leu Glu Gln Asp Ile Met Tyr Lys Gly Val Val Leu Arg		
	785	790	795
	800		
40	Tyr Ser Gly Ser Ser Ala Ala Leu Glu Arg Ile Arg Ser Phe Ser Pro		
	805	810	815
	Leu Lys Glu Pro Leu Thr Ile Gln Val Leu Thr Val Gly Asn Ala Leu		
	820	825	830
45	Arg Pro Lys Ile Lys Tyr Thr Tyr Phe Val Lys Lys Lys Glu Ser		
	835	840	845
	Phe Asn Ala Ile Pro Thr Phe Ser Ala Trp Val Ile Glu Glu Trp Gly		
	850	855	860
50	Glu Cys Ser Lys Ser Cys Glu Leu Gly Trp Gln Arg Arg Leu Val Glu		
	865	870	875
	880		
55	Cys Arg Asp Ile Asn Gly Gln Pro Ala Ser Glu Cys Ala Lys Glu Val		
	885	890	895
	Lys Pro Ala Ser Thr Arg Pro Cys Ala Asp His Pro Cys Pro Gln Trp		
	900	905	910
60	Gln Leu Gly Glu Trp Ser Ser Cys Ser Lys Thr Cys Gly Lys Gly Tyr		

915	920	925
Lys Lys Arg Ser Leu Lys Cys Leu Ser His Asp Gly Gly Val Leu Ser		
930	935	940
His Glu Ser Cys Asp Pro Leu Lys Pro Lys His Phe Ile Asp Phe		
945	950	955
Cys Thr Met Ala Glu Cys Ser		
10	965	

15 (2) INFORMATION FOR SEQ ID NO: 175:

(i) SEQUENCE CHARACTERISTICS:			
(A) LENGTH: 19 amino acids			
(B) TYPE: amino acid			
(D) TOPOLOGY: linear			
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 175:		

Met	Leu	Lys	Ile	Pro	Thr	His	Leu	Glu	Gly	Lys	Ile	Lys	Ile	Thr	Lys
1							5			10			15		

25 Val Tyr Xaa

30 (2) INFORMATION FOR SEQ ID NO: 176:

(i) SEQUENCE CHARACTERISTICS:			
(A) LENGTH: 205 amino acids			
(B) TYPE: amino acid			
(D) TOPOLOGY: linear			
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 176:		

Met	Tyr	Glu	Thr	Met	Lys	Leu	Asp	Ala	Cys	Xaa	His	Gln	Gln	Arg	Pro
1						5			10			15			
40	Thr	Leu	Gln	Ala	Gly	Pro	Lys	Leu	Leu	Thr	Leu	Ala	Pro	Arg	Glu
						20			25			30			

Pro	Arg	Gly	Gln	Ser	Gly	Arg	Gly	Ser	Glu	Leu	Thr	Ala	Arg	Gln	Arg	
35						40			45							
45	His	Ser	Thr	Gly	Asp	Pro	Gln	Gly	Glu	Gln	Ala	Leu	Pro	Arg	Ala	Gly
						50			55			60				

Leu	Leu	Arg	Thr	His	Pro	Asp	Ala	Arg	Pro	Lys	Ser	Ala	Met	Ala	Gln	
							85			90			95			
55	Cys	Val	Thr	Gly	Pro	Pro	Ala	Thr	Pro	His	Arg	Pro	Ser	Glu	Pro	Gln
							65			70			75		80	

Leu	Leu	Arg	Thr	His	Pro	Asp	Ala	Arg	Pro	Lys	Ser	Ala	Met	Ala	Gln	
							85			90			95			
60	Thr	Phe	Val	His	Gln	Gly	Pro	Val	Ala	Leu	Gln	Leu	Thr	Thr	Asn	
							100			105			110			
65	Arg	Arg	Val	Glu	Thr	Ser	Met	Ser	Ser	Asp	Gly	His	Gly	Gln	Asn	Pro
							115			120			125			

Thr Pro Ser Pro Trp Ala Asp Val Cys Ala Ser Arg Ala Asp Ala Val
 130 135 140
 5 Ala Phe Pro Ala Ser Gly Xaa Cys His Ser Pro Trp Leu Met Xaa Pro
 145 150 155 160
 Ser Ser His Pro Leu Asn Pro His Ser Pro Leu Asn Leu Pro Pro Pro
 165 170 175
 10 Ser Phe His Cys Lys Asp Pro Val Met Thr Leu His Pro Gln Thr Leu
 180 185 190
 Val Thr Gln Gly His Leu Ser Thr Ser Gly Arg Leu Thr
 15 195 200 205

20 (2) INFORMATION FOR SEQ ID NO: 177:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 54 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 177:
 Met Asp Ser Met Pro Glu Pro Ala Ser Arg Cys Leu Leu Leu Pro
 1 5 10 15
 30 Leu Leu Leu Leu Leu Leu Leu Pro Ala Pro Glu Leu Gly Pro
 20 25 30
 Ser Gin Ala Gly Ala Glu Glu Asn Asp Trp Val Arg Leu Pro Ser Lys
 35 40 45
 35 Cys Glu Gly Thr Cys Gly
 50

40 (2) INFORMATION FOR SEQ ID NO: 178:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 436 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 178:
 Met Pro Leu Phe Leu Leu Ser Leu Pro Thr Pro Pro Ser Ala Ser Gly
 50 1 5 10 15
 His Glu Arg Arg Gln Arg Pro Glu Ala Lys Thr Ser Gly Ser Glu Lys
 20 25 30
 55 Lys Tyr Leu Arg Ala Met Gln Ala Asn Arg Ser Gln Leu His Ser Pro
 35 40 45
 Pro Gly Thr Gly Ser Ser Glu Asp Ala Ser Thr Pro Gln Cys Val His
 50 55 60
 60

Thr Arg Leu Thr Gly Glu Gly Ser Cys Pro His Ser Gly Asp Val His
 65 70 75 80

5 Ile Gln Ile Asn Ser Ile Pro Lys Glu Cys Ala Glu Asn Ala Ser Ser
 85 90 95

Arg Asn Ile Arg Ser Gly Val His Ser Cys Ala His Gly Cys Val His
 100 105 110

10 Ser Arg Leu Arg Gly His Ser His Ser Glu Ala Arg Leu Thr Asp Asp
 115 120 125

Thr Ala Ala Glu Ser Gly Asp His Gly Ser Ser Ser Phe Ser Glu Phe
 130 135 140

15 Arg Tyr Leu Phe Lys Trp Leu Gln Lys Ser Leu Pro Tyr Ile Leu Ile
 145 150 155 160

Leu Ser Val Lys Leu Val Met Gln His Ile Thr Gly Ile Ser Leu Gly
 20 165 170 175

Ile Gly Leu Leu Thr Thr Phe Met Tyr Ala Asn Lys Ser Ile Val Asn
 180 185 190

25 Gln Val Phe Leu Arg Glu Arg Ser Ser Lys Ile Gln Cys Ala Trp Leu
 195 200 205

Leu Val Phe Leu Ala Gly Ser Ser Val Leu Leu Tyr Tyr Thr Phe His
 210 215 220

30 Ser Gln Ser Leu Tyr Tyr Ser Leu Ile Phe Leu Asn Pro Thr Leu Asp
 225 230 235 240

His Leu Ser Phe Trp Glu Val Phe Xaa Ile Val Gly Xaa Thr Asp Phe
 35 245 250 255

Ile Leu Lys Phe Phe Phe Met Gly Leu Lys Cys Leu Ile Leu Leu Val
 260 265 270

40 Pro Ser Phe Ile Met Pro Phe Lys Ser Lys Gly Tyr Trp Tyr Met Leu
 275 280 285

Leu Glu Glu Leu Cys Gln Tyr Tyr Arg Thr Phe Val Pro Ile Pro Val
 45 290 295 300

Trp Phe Arg Tyr Leu Ile Ser Tyr Gly Glu Phe Gly Xaa Val Thr Arg
 305 310 315 320

Trp Xaa Leu Gly Ile Leu Leu Ala Leu Leu Tyr Leu Ile Leu Lys Leu
 50 325 330 335

Leu Glu Phe Phe Gly His Leu Arg Thr Phe Arg Gln Val Leu Arg Ile
 340 345 350

55 Phe Phe Thr Xaa Pro Ser Tyr Gly Val Ala Ala Ser Lys Arg Gln Cys
 355 360 365

Ser Asp Val Asp Asp Ile Cys Ser Ile Cys Gln Ala Glu Phe Gln Lys
 370 375 380

60

Pro Ile Leu Leu Ile Cys Gln His Ile Phe Cys Glu Glu Cys Met Thr
 385 390 395 400

5 Leu Trp Phe Asn Arg Glu Lys Thr Cys Pro Leu Cys Arg Thr Val Ile
 405 410 415

Ser Asp His Ile Asn Lys Trp Lys Asp Gly Ala Thr Ser Ser His Leu
 420 425 430

10 Gln Ile Tyr Xaa
 435

15 (2) INFORMATION FOR SEQ ID NO: 179:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 175 amino acids
- (B) TYPE: amino acid

20 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

Val Val Phe Gly Ala Ser Leu Phe Leu Leu Ser Leu Thr Val Phe
 1 5 10 15

25 Ser Ile Val Ser Val Thr Ala Tyr Ile Ala Leu Ala Leu Ser Val
 20 25 30

30 Thr Ile Ser Phe Arg Ile Tyr Lys Gly Val Ile Gln Ala Ile Gln Lys
 35 40 45

Ser Asp Glu Gly His Pro Phe Arg Ala Tyr Leu Glu Ser Glu Val Ala
 50 55 60

35 Ile Ser Glu Glu Leu Val Gln Lys Tyr Ser Asn Ser Ala Leu Gly His
 65 70 75 80

Val Asn Cys Thr Ile Lys Glu Leu Arg Arg Leu Phe Leu Val Asp Asp
 85 90 95

40 Leu Val Asp Ser Leu Lys Phe Ala Val Leu Met Trp Val Phe Thr Tyr
 100 105 110

45 Val Gly Ala Leu Phe Asn Gly Leu Thr Leu Leu Ile Leu Ala Leu Ile
 115 120 125

Ser Leu Phe Ser Val Pro Val Ile Tyr Glu Arg His Gln Ala Gln Ile
 130 135 140

50 Asp His Tyr Leu Gly Leu Ala Asn Lys Asn Val Lys Asp Ala Met Ala
 145 150 155 160

Lys Ile Gln Ala Lys Ile Pro Gly Leu Lys Arg Lys Ala Glu Xaa
 165 170 175

55

(2) INFORMATION FOR SEQ ID NO: 180:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 219 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

5

Met	Glu	Ala	Pro	Gly	Ala	Pro	Pro	Arg	Thr	Leu	Thr	Trp	Glu	Ala	Met
1	5													15	

10

Glu	Gln	Ile	Arg	Tyr	Leu	His	Glu	Glu	Phe	Pro	Glu	Trp	Ser	Val
	20				25						30			

15

Pro	Arg	Leu	Ala	Glu	Gly	Phe	Asp	Val	Ser	Thr	Asp	Val	Ile	Arg	Arg
		35				40						45			

20

Val	Leu	Lys	Ser	Lys	Phe	Leu	Pro	Thr	Leu	Glu	Gln	Lys	Leu	Lys	Gln
	50				55				60						
Asp	Gln	Lys	Val	Leu	Lys	Lys	Ala	Gly	Leu	Ala	His	Ser	Leu	Gln	His
	65				70				75				80		

25

Leu	Arg	Gly	Ser	Gly	Asn	Thr	Ser	Lys	Leu	Leu	Pro	Ala	Gly	His	Ser
	85					90						95			

30

Val	Ser	Gly	Ser	Leu	Leu	Met	Pro	Gly	His	Glu	Ala	Ser	Ser	Lys	Asp
	100					105				110					
Pro	Asn	His	Ser	Thr	Ala	Leu	Lys	Val	Ile	Glu	Ser	Asp	Thr	His	Arg
	115					120				125					

35

Thr	Asn	Thr	Pro	Arg	Arg	Lys	Gly	Arg	Asn	Lys	Glu	Ile	Gln	Asp	
	130				135				140						
Leu	Glu	Glu	Ser	Phe	Val	Pro	Val	Ala	Ala	Pro	Leu	Gly	His	Pro	Arg
	145				150				155			160			

40

Glu	Leu	Gln	Lys	Tyr	Ser	Ser	Asp	Ser	Glu	Ser	Pro	Arg	Gly	Thr	Gly
	165				170				175						
Ser	Gly	Ala	Leu	Pro	Ser	Gly	Gln	Lys	Leu	Glu	Leu	Lys	Ala	Glu	
	180				185				190						

45

Glu	Pro	Asp	Asn	Phe	Ser	Ser	Lys	Val	Val	Gln	Arg	Gly	Arg	Glu	Phe
	195				200				205						
Phe	Asp	Ser	Asn	Gly	Asn	Phe	Leu	Tyr	Arg	Ile					
	210				215										

50

(2) INFORMATION FOR SEQ ID NO: 181:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

Trp

Lys

1

Ala

5

Glu

Leu

Xaa

60

(2) INFORMATION FOR SEQ ID NO: 182:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

10

Met Ser Asn Thr Leu Leu Ser Gln Trp Leu Leu Leu Leu Thr Leu Phe
1 5 10 15

15

Lys Cys Ile Ile Leu Pro Leu Asn Leu Xaa Pro Ile Ile Arg Thr Ile
20 25 30

Pro Asp Trp Ser Pro Glu Leu Gly Thr Asn Thr Xaa
35 40

20

(2) INFORMATION FOR SEQ ID NO: 183:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 59 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

30

Met Trp Gln Val Arg Arg Gly Gly Cys Val Leu Ala Val Cys Ser Gln
1 5 10 15

Ala Arg Gly Thr Gly Gly Arg Leu Gly Trp Val Gly Thr Ser Ser Leu
20 25 30

35

Arg Val Arg Met Ala Glu Ser Thr Ser Leu Met Ser Gln Gly Arg Ser
35 40 45

40

Pro Ile Pro Arg Met Thr Pro Ala Arg Pro Xaa
50 55

45

(2) INFORMATION FOR SEQ ID NO: 184:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 588 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

Met Arg Asp Ala Gly Asp Pro Ser Pro Pro Asn Lys Met Leu Arg Arg
1 5 10 15

55

Ser Asp Ser Pro Glu Asn Lys Tyr Ser Asp Ser Thr Gly His Ser Lys
20 25 30

Ala Lys Asn Val His Thr His Arg Val Arg Glu Arg Asp Gly Gly Thr
35 40 45

60

Ser Tyr Ser Pro Gln Glu Asn Ser His Asn His Ser Ala Leu His Ser
 50 55 60

5 Ser Asn Ser His Ser Ser Asn Pro Ser Asn Asn Pro Ser Lys Thr Ser
 65 70 75 80

Asp Ala Pro Tyr Asp Ser Ala Asp Asp Trp Ser Glu His Ile Ser Ser
 85 90 95

10 Ser Gly Lys Lys Tyr Tyr Asn Cys Arg Thr Glu Val Ser Gln Trp
 100 105 110

Glu Lys Pro Lys Glu Trp Leu Glu Arg Glu Gln Arg Gln Lys Glu Ala
 115 120 125

15 Asn Lys Met Ala Val Asn Ser Phe Pro Lys Asp Arg Asp Tyr Arg Arg
 130 135 140

Glu Val Met Gln Ala Thr Ala Thr Ser Gly Phe Ala Ser Gly Met Glu
 20 145 150 155 160

Asp Lys His Ser Ser Asp Ala Ser Ser Leu Leu Pro Gln Asn Ile Leu
 165 170 175

25 Ser Gln Thr Ser Arg His Asn Asp Arg Asp Tyr Arg Leu Pro Arg Ala
 180 185 190

Glu Thr His Ser Ser Ser Thr Pro Val Gln His Pro Ile Lys Pro Val
 30 195 200 205

Val His Pro Thr Ala Thr Pro Ser Thr Val Pro Ser Ser Pro Phe Thr
 210 215 220

35 Leu Gln Ser Asp His Gln Pro Lys Lys Ser Phe Asp Ala Asn Gly Ala
 225 230 235 240

Ser Thr Leu Ser Lys Leu Pro Thr Pro Thr Ser Ser Val Pro Ala Gln
 245 250 255

40 Lys Thr Glu Arg Lys Glu Ser Thr Ser Gly Asp Lys Pro Val Ser His
 260 265 270

Ser Cys Thr Thr Pro Ser Thr Ser Ser Ala Ser Gly Leu Asn Pro Thr
 45 275 280 285

Ser Ala Pro Pro Thr Ser Ala Ser Ala Val Pro Val Ser Pro Val Pro
 290 295 300

Gln Ser Pro Ile Pro Pro Leu Leu Gln Asp Pro Asn Leu Leu Arg Gln
 50 305 310 315 320

Leu Leu Pro Ala Leu Gln Ala Thr Leu Gln Leu Asn Asn Ser Asn Val
 325 330 335

55 Asp Ile Ser Lys Ile Asn Glu Val Leu Thr Ala Ala Val Thr Gln Ala
 340 345 350

Ser Leu Gln Ser Ile Ile His Lys Phe Leu Thr Ala Gly Pro Ser Ala
 60 355 360 365

Phe Asn Ile Thr Ser Leu Ile Ser Gln Ala Ala Gln Leu Ser Thr Gln
 370 375 380
 Ala Gln Pro Ser Asn Gln Ser Pro Met Ser Leu Thr Ser Asp Ala Ser
 5 385 390 395 400
 Ser Pro Arg Ser Tyr Val Ser Pro Arg Ile Ser Thr Pro Gln Thr Asn
 405 410 415
 10 Thr Val Pro Ile Lys Pro Leu Ile Ser Thr Pro Pro Val Ser Ser Gln
 420 425 430
 Pro Lys Val Ser Thr Pro Val Val Lys Gln Gly Pro Val Ser Gln Ser
 435 440 445
 15 Ala Thr Gln Gln Pro Val Thr Ala Asp Lys Xaa Gln Gly His Glu Pro
 450 455 460
 Val Ser Pro Arg Ser Leu Gln Arg Ser Ser Ser Gln Arg Ser Pro Ser
 20 465 470 475 480
 Pro Gly Pro Asn His Thr Ser Asn Ser Ser Asn Ala Ser Asn Ala Thr
 485 490 495
 25 Val Val Pro Gln Asn Ser Ser Ala Arg Ser Thr Cys Ser Leu Thr Pro
 500 505 510
 Ala Leu Ala Ala His Phe Ser Glu Asn Leu Ile Lys His Val Gln Gly
 515 520 525
 30 Trp Pro Ala Asp His Ala Glu Lys Gln Ala Ser Arg Leu Arg Glu Glu
 530 535 540
 Ala His Asn Met Gly Thr Ile His Met Ser Glu Ile Cys Thr Glu Leu
 35 545 550 555 560
 Lys Asn Leu Arg Ser Leu Val Arg Val Cys Glu Ile Gln Ala Thr Leu
 565 570 575
 40 Arg Glu Gln Arg Asp Thr Ile Phe Glu Thr Thr Asn
 580 585

45 (2) INFORMATION FOR SEQ ID NO: 185:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 166 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 185:
 Met Asn Ile Lys His Leu Val Asp Pro Ile Asp Asp Leu Phe Leu Ala
 1 5 10 15
 55 Ala Lys Lys Ile Pro Gly Ile Ser Ser Thr Gly Val Gly Asp Gly Gly
 20 25 30
 Asn Glu Leu Gly Met Gly Lys Val Lys Glu Ala Val Arg Arg His Ile
 60 35 40 45

Arg His Gly Asp Val Ile Ala Cys Asp Val Glu Ala Asp Phe Ala Val
 50 55 60

5 Ile Ala Gly Val Ser Asn Trp Gly Gly Tyr Ala Leu Ala Cys Ala Leu
 65 70 75 80

Tyr Ile Leu Tyr Ser Cys Ala Val His Ser Gln Tyr Leu Arg Lys Ala
 85 90 95

10 Val Gly Pro Ser Arg Ala Pro Gly Asp Gln Ala Trp Thr Gln Ala Leu
 100 105 110

15 Pro Ser Val Ile Lys Glu Glu Lys Met Leu Gly Ile Leu Val Gln His
 115 120 125

Lys Val Arg Ser Gly Val Ser Gly Ile Val Gly Met Glu Val Asp Gly
 130 135 140

20 Leu Pro Phe His Asn Xaa His Ala Glu Met Ile Gln Lys Leu Val Asp
 145 150 155 160

Val Thr Thr Ala Gln Val
 165

25

(2) INFORMATION FOR SEQ ID NO: 186:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 9 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

35 Met Leu Ile Leu Phe Leu Lys Lys Xaa
 1 5

40 (2) INFORMATION FOR SEQ ID NO: 187:
 (i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 20 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

50 Thr His Thr His Pro Lys Ser Phe Tyr Ile Ile Lys Leu Ser
 1 5 10 15

Tyr Tyr Tyr Xaa
 20

55

(2) INFORMATION FOR SEQ ID NO: 188:

60 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

5 Met Ile Gln Ser Gly Leu Ile Ala Ile Leu Leu Ser Phe Leu Lys Val
1 5 10 15

Tyr Val Glu Gly Arg Pro Cys Val Cys Phe Ser Lys Gly Leu Xaa Xaa
20 25 30

10

15

(2) INFORMATION FOR SEQ ID NO: 189:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

20 (B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

Tyr Ile Tyr Leu Ile Val Tyr Ile Ser Phe Tyr Ser Phe Arg Pro Gln
25 1 5 10 15

Gln Leu Xaa

30

(2) INFORMATION FOR SEQ ID NO: 190:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

40 Met Arg Phe Leu Leu Thr Val Trp Gly Ser Phe Pro Phe Met Leu Ile
1 5 10 15

Pro Val Phe Leu Ser Ile Gly Thr Lys Glu Met Lys Lys Ala Gln Arg
20 25 30

45

Xaa

50

(2) INFORMATION FOR SEQ ID NO: 191:

(i) SEQUENCE CHARACTERISTICS:

55 (A) LENGTH: 84 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

60 Met Arg Val Pro Pro Val Leu Arg Gly Arg Ile Leu Pro Leu Val Leu
1 5 10 15

20 Gln Cys Thr Leu Leu Glu Phe Cys Leu Cys Ala Thr Thr Val Leu Pro
25 30

5 Thr Val Xaa Cys Trp Lys Pro Arg Leu Pro Val Xaa Ala Ser Gly Leu
35 40 45

10 Tyr Val Asp Arg Met Ser Leu Trp Lys Tyr Gly Cys Ser Gly Trp Asn
50 55 60

15 Glu Ser Ala Arg Pro Arg Arg Ala Gly Gly Thr Met Arg Pro Pro Arg
65 70 75 80

Ser Gly Arg Xaa

20 (2) INFORMATION FOR SEQ ID NO: 192:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 123 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

Met Ala Gly Ala Phe Val Ala Val Phe Leu Leu Ala Met Phe Tyr Glu
1 5 10 15

30 Gly Leu Lys Ile Ala Arg Glu Ser Leu Leu Arg Lys Ser Gln Val Ser
20 25 30

Ile Arg Tyr Asn Ser Met Pro Val Pro Gly Pro Asn Gly Thr Ile Leu
35 40 45

35 Met Glu Thr His Lys Thr Val Gly Gln Gln Met Leu Ser Phe Pro His
50 55 60

40 Leu Leu Gln Thr Val Leu His Ile Ile Gln Val Val Ile Ser Tyr Phe
65 70 75 80

Leu Met Leu Ile Phe Met Thr Tyr Asn Gly Tyr Leu Cys Ile Ala Xaa
85 90 95

45 Ala Ala Gly Ala Gly Thr Gly Tyr Phe Leu Phe Ser Trp Lys Lys Ala
100 105 110

50 Val Val Val Asp Ile Thr Glu His Cys His Xaa
115 120

55 (2) INFORMATION FOR SEQ ID NO: 193:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 143 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

Met Gly Cys Leu Val Trp Gly Pro Ser Trp Pro Pro Leu Ser Leu Leu
 1 5 10 15

Ala Ser Leu Leu His Ser Gly Ile Ala Gly Arg Cys Leu Leu Cys Leu
 5 20 25 30

Phe Lys Gly Leu Ala Ala Ala Ser Leu Gln Ile Arg Asp Leu Ala
 35 40 45

10 Ser Arg Leu Thr Thr Gly Pro Arg Thr Cys Arg Val Gln Pro Pro Pro
 50 55 60

His Pro Gln Ser Ser Pro Pro Trp Pro Gly Pro Pro Gly Ala Glu Thr
 65 70 75 80

15 Cys Arg Pro Leu Ser Arg Thr Val Gly Gly Val Cys Pro Ser Asp Trp
 85 90 95

Pro Val Ser Trp Leu Leu Leu Pro Pro Leu Pro Glu Val Val Thr Cys
 20 100 105 110

Ser Cys Pro Arg Ile Lys Ala Arg Pro Glu Arg Thr Pro Glu Leu Leu
 115 120 125

25 Cys Ala Trp Gly Gly Arg Gly Lys His Ser Gln Leu Val Ala Xaa
 130 135 140

30 (2) INFORMATION FOR SEQ ID NO: 194:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 amino acids
 (B) TYPE: amino acid
 35 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

 Met Pro Asn Val Met Leu Thr Leu Phe Val Met Thr Leu Ser Ser Ala
 1 5 10 15

40 Ser Asn Leu Gly Leu Tyr Phe Phe Lys Phe Asn Phe Glu Cys Ser Cys
 20 25 30

Met Phe Gly Thr Ser Leu Leu Thr Ala Lys Asp Lys Leu Phe Ile Cys
 45 35 40 45

Ile Thr Xaa
 50 50

50 (2) INFORMATION FOR SEQ ID NO: 195:

(i) SEQUENCE CHARACTERISTICS:
 55 (A) LENGTH: 222 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 195:

 60 Met Ser Leu Leu Val Leu Val Ser Trp Gly Ser Met Gly Leu Glu

	1	5	10	15
	Ala Ala Thr Ala Val Gly Leu Ser Asp Phe Cys Ser Asn Pro Asp Pro			
	20	25	30	
5	Tyr Val Leu Asn Leu Thr Gln Glu Glu Thr Gly Leu Ser Ser Asp Ile			
	35	40	45	
10	Leu Ser Tyr Tyr Leu Leu Cys Asn Arg Ala Val Ser Asn Pro Phe Gln			
	50	55	60	
	Gln Arg Leu Thr Leu Ser Gln Arg Ala Leu Ala Asn Ile His Ser Gln			
	65	70	75	80
15	Leu Leu Gly Leu Glu Arg Glu Ala Val Pro Gln Phe Pro Ser Ala Gln			
	85	90	95	
	Lys Pro Leu Leu Ser Leu Glu Glu Thr Leu Asn Val Thr Glu Gly Asn			
	100	105	110	
20	Phe His Gln Leu Val Ala Leu Leu His Cys Arg Ser Leu His Lys Asp			
	115	120	125	
25	Tyr Gly Ala Ala Leu Arg Gly Leu Cys Glu Xaa Xaa Leu Glu Gly Leu			
	130	135	140	
	Leu Phe Leu Leu Leu Phe Ser Leu Leu Ser Ala Gly Ala Leu Ala Xaa			
	145	150	155	160
30	Ala Leu Cys Xaa Leu Pro Arg Ala Trp Ala Leu Phe Pro Pro Arg Asn			
	165	170	175	
	Pro Ser Ala Leu Cys Ser Gly Ser Arg Leu Ser Glu Pro Leu Leu Pro			
	180	185	190	
35	Ala Gly Leu Glu Pro Gly Ser Pro Leu Arg Ser Phe Pro Gly Cys Arg			
	195	200	205	
40	Arg Asp Pro Thr Asn Pro Ala Cys Leu Gly Ser Asp His Xaa			
	210	215	220	

(2) INFORMATION FOR SEQ ID NO: 196:

45	(i) SEQUENCE CHARACTERISTICS:			
	(A) LENGTH: 102 amino acids			
	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 196:			

Met Ser Gln Leu Ser Arg Thr Ser Leu Ser Leu Leu Leu Thr Leu Leu			
1	5	10	15

55 Val Leu Trp Gly Ser Ser Cys Cys Leu Pro Ile Trp Cys Leu Pro Asn			
20	25	30	

Arg His Arg Leu Leu Lys Leu Ser Phe Leu Leu Phe Ser Pro Asp Ile			
35	40	45	

Pro Tyr Leu Ser His Thr His Pro Asn Asn Ile Ser Cys Ser Val Leu
50 55 60

5 Ser Leu Arg Gln His Leu Asn Phe Thr Gln Pro Gly Ala Leu Phe Thr
65 70 75 80

Cys Leu Val Gln Ile Gln Phe Gly Leu Ile Leu Gln Pro Cys Ile Ser
85 90 95

10 Lys Trp Gly Leu Gly Xaa
100

15 (2) INFORMATION FOR SEQ ID NO: 197:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
20 (D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

Met Ile Ala Leu Phe Phe Val Thr Thr Xaa Leu Thr Xaa
1 5 10

25

(2) INFORMATION FOR SEQ ID NO: 198:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 61 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

Met Thr Tyr His Pro Asn Gln Val Val Glu Gly Cys Cys Ser Asp Met
1 5 10 15

40 Ala Val Thr Phe Asn Gly Leu Thr Pro Asn Gln Met His Val Met Met
20 25 30

Tyr Gly Val Tyr Arg Leu Arg Ala Phe Gly His Ile Phe Asn Asp Ala
35 40 45

45 Leu Val Phe Leu Pro Pro Asn Gly Ser Asp Asn Asp Xaa
50 55 60

50 (2) INFORMATION FOR SEQ ID NO: 199:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
55 (D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

Met Ser Ser Ser Ser Leu His Trp Lys Glu Phe Lys Tyr Ala Pro Gly
1 5 10 15

60

Ser Leu His Tyr Phe Ala Leu Ser Phe Val Leu Ile Leu Thr Glu Ile
20 25 30

5 Cys Leu Val Ser Ser Gly Met Gly Phe Pro Gln Glu Gly Lys His Phe
35 40 45

Ser Val Leu Gly Ser Pro Asp Cys Ser Leu Trp Gly Arg Asp Glu His
50 55 60

10 Val Pro Arg Glu Phe Ala Xaa
65 70

15 (2) INFORMATION FOR SEQ ID NO: 200:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 amino acids
(B) TYPE: amino acid
20 (D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

Met His Leu Arg Phe Pro Phe Leu Cys Xaa
1 5 10
25

(2) INFORMATION FOR SEQ ID NO: 201:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

Met Arg Arg Val Ala Arg Gly Arg Gly Leu Ala Leu Pro Ser Leu Glu
1 5 10 15

40 His Arg Pro Ser Cys Ser Tyr Asp Ala Leu Pro Leu Pro Phe Cys Glu
20 25 30

Thr Arg Asn Pro Glu Ala His Leu Tyr Phe Phe Arg Thr Asp Val Glu
35 40 45

45 Arg Xaa
50

50 (2) INFORMATION FOR SEQ ID NO: 202:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 amino acids
(B) TYPE: amino acid
55 (D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

Ala Lys Ile Leu Val Phe Ile Phe Leu Phe Glu Leu Xaa
1 5 10
60

(2) INFORMATION FOR SEQ ID NO: 203:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

10 Met Phe Gln Glu Cys Ile Pro Ile Ser Leu Phe Phe Leu Asn Trp Leu
1 5 10 15

15 Lys Glu Cys Cys Ser Phe Thr Cys Pro Asn Ser His Ile Asn Asn Cys
20 25 30

Leu Thr Gly Ile Arg Xaa
35

20

(2) INFORMATION FOR SEQ ID NO: 204:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

30 Met Asn Phe Val Leu Phe Phe Ile Gly Ile Asn Val Gly Cys Arg Gly
1 5 10 15

Glu Asn Ser Leu Lys Tyr Phe Thr Val Thr Val Xaa Cys Ser Pro Arg
20 25 30

35 Asp Xaa

40

(2) INFORMATION FOR SEQ ID NO: 205:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

50 Met Leu Leu Phe Leu Phe Val Cys Leu Pro Ile Thr Trp Met Ala Glu
1 5 10 15

Phe Leu Ser Gln Leu Arg His Leu Leu Xaa
20 25

55

(2) INFORMATION FOR SEQ ID NO: 206:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 105 amino acids

60

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

5 Met Pro Arg His Ser Leu Tyr Ile Ile Ile Gly Ala Leu Cys Val Ala
 1 5 10 15

Phe Ile Leu Met Leu Ile Ile Leu Ile Val Gly Ile Cys Arg Ile Ser
 20 25 30

10 Arg Ile Glu Tyr Gln Gly Ser Ser Arg Pro Ala Tyr Glu Glu Phe Tyr
 35 40 45

15 Asn Cys Arg Ser Ile Asp Ser Glu Phe Ser Asn Ala Ile Ala Ser Ile
 50 55 60

Arg His Ala Arg Phe Gly Lys Lys Ser Arg Pro Ala Met Tyr Asp Val
 65 70 75 80

20 Ser Pro Ile Ala Tyr Glu Asp Tyr Ser Pro Asp Asp Lys Pro Leu Val
 85 90 95

Thr Leu Ile Lys Thr Lys Asp Leu Xaa
 100 105

25

(2) INFORMATION FOR SEQ ID NO: 207:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 64 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

35 Leu Lys Ser Cys Leu Leu Leu Val Ser Phe Leu Ser Gly Arg Val Pro
 1 5 10 15

40 Ser Tyr Asp Leu Ile Tyr Val Cys Ser Ile Ala Leu Glu Thr Gly Phe
 20 25 30

Val Cys Glu Met Ala Leu Ser Phe Val Asp His Phe Cys Arg Glu Ile
 35 40 45

45 Val Asp Leu Gly Arg Ala Glu Ala Thr Ala Asp Met Pro Gly Val Xaa
 50 55 60

50

(2) INFORMATION FOR SEQ ID NO: 208:

55 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 42 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

60

Met Ser Ala Trp Leu Pro Ser Pro Pro His Leu Leu Leu Leu Ser Ala
1 5 10 15

5 Ala Ala Gly Ser Gly Ala Ser His Leu Arg Ala Leu Gly Ser Ser Ala
20 25 30

Leu Glu Gly Leu Gln Asp Pro Ser Gln Xaa
35 40

10

(2) INFORMATION FOR SEQ ID NO: 209:

15 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

20 Met Ser Ser Pro Ala Thr Trp Arg Leu Thr Leu Pro Ser Leu Leu Val
1 5 10 15

Phe Leu Thr Gly Glu Ala Met Pro Trp Pro Ala His Ser Thr Ser Cys
20 25 30

25

Thr His Val Leu Ser Thr Val Ser Thr Xaa
35 40

30

(2) INFORMATION FOR SEQ ID NO: 210:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

40 Met Gln Ala Pro Leu Gln Asp Cys Gly Arg Ser Val Ser Leu Arg Leu
1 5 10 15

Ala Cys Val Leu Ala Pro Leu Thr Thr Ser Ser Arg Gly Cys His Leu
20 25 30

45

Gln Leu Pro Gln Asp Lys Gly Lys Ala Arg Xaa Asp Ser Xaa
35 40 45

50

(2) INFORMATION FOR SEQ ID NO: 211:

55 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 266 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 211:

Met Asn Gly Ser His Lys Asp Pro Leu Leu Pro Phe Pro Ala Ser Ala
1 5 10 15

60

Arg Thr Pro Ser Leu Pro Pro Ala Pro Pro Ala Gln Ala Pro Leu Pro
 20 25 30

5 Trp Lys Pro Ser Gly Phe Ala Arg Ile Ser Pro Pro Pro Pro Leu Ala
 35 40 45

Ile Leu Gln Tyr Arg Gly Lys Ala Asp His Gly Glu Ser Gly Gln Gln
 50 55 60

10 Leu Ala Ala Ala Pro Gly Asp Gly Arg Leu Pro Leu Leu Glu Ala Val
 65 70 75 80

Arg Arg Leu Arg Gly Gln Asp Cys Gly Pro Leu Ser Ala Leu Cys His
 85 90 95

15 Gly Gln Leu Leu Ala Gln Pro Val Pro Gln Val Leu Leu Leu Pro Gly
 100 105 110

Ala Xaa Gly Asp Ile Gly Thr Ser Cys Tyr Thr Lys Ser Gly Met Ile
 20 115 120 125

Leu Cys Arg Asn Asp Tyr Ile Arg Leu Phe Gly Asn Ser Gly Ala Cys
 130 135 140

25 Ser Ala Cys Gly Gln Ser Ile Pro Ala Ser Glu Leu Val Met Arg Ala
 145 150 155 160

Gln Gly Asn Val Tyr His Leu Lys Cys Phe Thr Cys Ser Thr Cys Arg
 165 170 175

30 Asn Arg Leu Val Pro Gly Asp Arg Phe His Tyr Ile Asn Gly Ser Leu
 180 185 190

Phe Cys Glu His Asp Arg Pro Thr Ala Leu Ile Asn Gly His Leu Asn
 35 195 200 205

Ser Leu Gln Ser Asn Pro Leu Leu Pro Asp Gln Lys Val Cys Lys Val
 210 215 220

40 Arg Val Met Gln Asn Ala Cys Leu His Leu Arg Phe Val His His Arg
 225 230 235 240

Trp Ile Pro Cys Xaa Phe Ser Arg Gln Val Thr Phe Val Ala Ser Thr
 45 245 250 255

Ser Ala Ser Ser Met Pro Leu His Leu Leu
 260 265

50 (2) INFORMATION FOR SEQ ID NO: 212:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 94 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 212:

Met Ala Arg Thr Arg Thr Pro Ser Ser Pro Phe Leu Leu Leu Arg Glu
 60 1 5 10 15

Leu Pro Pro Ser Leu Gln Leu Arg Gln Pro Arg Arg Pro Phe Pro Gly
 20 25 30

5 Ser Arg Ala Ala Ser Leu Ala Phe His Arg Arg Arg Leu Ser Gln Tyr
 35 40 45

Cys Asn Ile Gly Glu Lys Gln Thr Met Val Asn Pro Gly Ser Ser Ser
 10 50 55 60

Gln Pro Pro Pro Val Thr Ala Gly Ser Leu Ser Trp Lys Arg Cys Ala
 65 70 75 80

15 Gly Cys Gly Lys Ile Ala Asp Arg Phe Leu Leu Tyr Ala
 85 90

20 (2) INFORMATION FOR SEQ ID NO: 213:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

Leu Phe Gly Asn Ser Gly Ala Cys Ser Ala Cys Gly Gln Ser Ile Pro
 1 5 10 15

30 Ala Ser Glu Leu Val Met Arg Ala
 20

35 (2) INFORMATION FOR SEQ ID NO: 214:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

His Asp Arg Pro Thr Ala Leu Ile Asn Gly His Leu Asn Ser Leu Gln
 1 5 10 15

45 Ser Asn Pro

50 (2) INFORMATION FOR SEQ ID NO: 215:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

60 Leu Val Pro Gly Asp Arg Phe His Tyr Ile Asn Gly
 1 5 10

5 (2) INFORMATION FOR SEQ ID NO: 216:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 81 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

Met Lys Tyr Met Gly Gly Cys Ala Lys Val Met Cys Lys Tyr Tyr Val
1 5 10 15

15 Ile Leu Tyr Gln Gly Leu Glu Tyr Pro Leu Leu Xaa Ser Gly Asp Pro
20 25 30

Glu Thr Ser Pro Pro Trp Ile Leu Arg Ala Asp Cys Ile Val Leu Ser
35 40 45

20 Ser Arg Asn Phe His Ser Asn Xaa Gly Arg Leu Thr Ile Asn Lys Ile
50 55 60

25 Tyr Val Ile Gly Gly Lys Tyr Arg Gly Glu Val Thr Asn Gly Ala
65 70 75 80

Lys

30

(2) INFORMATION FOR SEQ ID NO: 217:

35 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

40 Met Gly Gln Ser Glu Leu Tyr Ser Ser Ile Leu Arg Asn Leu Gly Val
1 5 10 15

Leu Phe Leu Val Tyr Thr Arg Gly Gly Phe Leu Leu Ser Pro Leu Leu
20 25 30

45 His Gly Thr Leu Thr Cys Ala His Ser
35 40

50

(2) INFORMATION FOR SEQ ID NO: 218:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 218:

60 Met Val Leu Leu Leu Thr Val Ala Ser Tyr Thr Val Phe Trp Met
1 5 10 15

Ile Gly Asp Val Leu Asp Ile Leu Phe Leu Trp Asn Phe Glu Tyr Thr
20 25 30

5 Thr Leu Tyr
35

10 (2) INFORMATION FOR SEQ ID NO: 219:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 amino acids
(B) TYPE: amino acid
15 (D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

Met Glu Leu Tyr Asn Ser Leu Cys Pro Ile Cys Tyr Phe Ser Thr Val
1 5 10 15

20 Leu Thr Thr Thr Tyr Ile Tyr Phe Val Tyr Ser Gln Ser Ser Xaa
20 25 30

25 Ile Arg Met Lys Val Pro
35

30 (2) INFORMATION FOR SEQ ID NO: 220:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids
(B) TYPE: amino acid
35 (D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

Met Gln Ile Val Ile Val Leu Tyr Cys Val Arg Asn Lys Asp Lys Lys
1 5 10 15

40 Lys Val Cys Thr Cys Ser Val Gln Thr Gln Phe Phe Phe Pro Ile Phe
20 25 30

45 Pro Ile Leu Gly Cys Leu Asn Gly Cys Arg Thr Gln Glu
35 40 45

(2) INFORMATION FOR SEQ ID NO: 221:

50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

Met Lys Tyr Met Gly Gly Cys Ala Lys Val Met Cys Lys Tyr Tyr Val
1 5 10 15

60 Ile Leu Tyr Gln Gly Leu Glu Tyr Pro Leu Leu Xaa
20 25

5 (2) INFORMATION FOR SEQ ID NO: 222:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

Leu Glu Tyr Pro Leu Leu Xaa Ser Gly Asp Pro Glu Thr Ser Pro Pro
1 5 10 15

15 Trp Ile Leu Arg Ala Asp Cys Ile Val Leu Ser Ser Arg Asn Phe His
20 25 30

Ser Asn Xaa
35

20

(2) INFORMATION FOR SEQ ID NO: 223:

25 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

Arg Asn Phe His Ser Asn Xaa Gly Arg Leu Thr Ile Asn Lys Ile Tyr
1 5 10 15

35 Val Ile Gly Gly Lys Tyr Arg Gly Glu Val Thr Asn Gly Ala Lys
20 25 30

40

(2) INFORMATION FOR SEQ ID NO: 224:

45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 145 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

50 Val Thr Asn Glu Met Ser Gln Gly Arg Gly Lys Tyr Asp Phe Tyr Ile
1 5 10 15

Gly Leu Gly Leu Ala Met Ser Ser Ser Ile Phe Ile Gly Gly Ser Phe
20 25 30

55 Ile Leu Lys Lys Lys Gly Leu Leu Arg Leu Ala Arg Lys Gly Ser Met
35 40 45

60 Arg Ala Gly Gln Gly His Ala Tyr Leu Lys Glu Trp Leu Trp Trp
50 55 60

Ala Gly Leu Leu Ser Met Gly Ala Gly Glu Val Ala Asn Phe Ala Ala
 65 70 75 80

5 Tyr Ala Phe Ala Pro Ala Thr Leu Val Thr Pro Leu Gly Ala Leu Ser
 85 90 95

Val Leu Val Ser Ala Ile Leu Ser Ser Tyr Phe Leu Asn Glu Arg Leu
 100 105 110

10 Asn Leu His Gly Lys Ile Gly Cys Leu Leu Ser Ile Leu Gly Ser Thr
 115 120 125

15 Val Met Val Ile His Ala Pro Lys Glu Glu Glu Ile Glu Thr Leu Asn
 130 135 140

Glu
 145

20

(2) INFORMATION FOR SEQ ID NO: 225:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 78 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

30 Val Thr Asn Glu Met Ser Gln Gly Arg Gly Lys Tyr Asp Phe Tyr Ile
 1 5 10 15

Gly Leu Gly Leu Ala Met Ser Ser Ser Ile Phe Ile Gly Gly Ser Phe
 20 25 30

35 Ile Leu Lys Lys Lys Gly Leu Leu Arg Leu Ala Arg Lys Gly Ser Met
 35 40 45

40 Arg Ala Gly Gln Gly Gly His Ala Tyr Leu Lys Glu Trp Leu Trp Trp
 50 55 60

Ala Gly Leu Leu Ser Met Gly Ala Gly Glu Val Ala Asn Phe
 65 70 75

45

(2) INFORMATION FOR SEQ ID NO: 226:

50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 30 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

55 Asn Phe Ala Ala Tyr Ala Phe Ala Pro Ala Thr Leu Val Thr Pro Leu
 1 5 10 15

Gly Ala Leu Ser Val Leu Val Ser Ala Ile Leu Ser Ser Tyr
 20 25 30

60

(2) INFORMATION FOR SEQ ID NO: 227:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

10

Glu Arg Leu Asn Leu His Gly Lys Ile Gly Cys Leu Leu Ser Ile Leu
1 5 10 15

15

Gly Ser Thr Val Met Val Ile His Ala Pro Lys Glu Glu Glu Ile Glu
20 25 30

Thr Leu Asn Glu
35

20

(2) INFORMATION FOR SEQ ID NO: 228:

25 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

30

Arg Phe Lys Thr Leu Met Thr Asn Lys Ser Glu Gln Asp Gly Asp Ser
1 5 10 15

35

Ser Lys Thr Ile Glu Ile Ser Asp Met Lys Tyr His Ile Phe Gln
20 25 30

(2) INFORMATION FOR SEQ ID NO: 229:

40 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

45

Leu Val Glu Gly Lys Leu Phe Tyr Ala His Lys Val Leu Leu Val Thr
1 5 10 15

50

Xaa Ser Asn Arg
20

(2) INFORMATION FOR SEQ ID NO: 230:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 87 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

5 CCTTAAAAGC TGACATTTTA TAATTGTGTT GTATAGCAGC AACTATATCC TTCCAAAAAT 60
 CAAATGTTT TTGACCATTG TTTCAGTT 87

10

(2) INFORMATION FOR SEQ ID NO: 231:

15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 38 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

CCCTAAAAGC TGACATTTTA TAATTGTGTT GTATAGCA 38

25

(2) INFORMATION FOR SEQ ID NO: 232:

30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 38 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

CTTCCAAAAA TCAAATGTTT TTGACCATT GTTCAGTT 38

40

(2) INFORMATION FOR SEQ ID NO: 233:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 455 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

50 Met Ala Gln His Phe Ser Leu Ala Ala Cys Asp Val Val Gly Phe Asp
 1 5 10 15

Leu Asp His Thr Leu Cys Arg Tyr Asn Leu Pro Glu Ser Ala Pro Leu
 20 25 30

55 Ile Tyr Asn Ser Phe Ala Gln Phe Leu Val Lys Glu Lys Gly Tyr Asp
 35 40 45

Lys Glu Leu Leu Asn Val Thr Pro Glu Asp Trp Asp Phe Cys Cys Lys
 50 55 60

Gly Leu Ala Leu Asp Leu Glu Asp Gly Asn Phe Leu Lys Leu Ala Asn
 65 70 75 80

5 Asn Gly Thr Val Leu Arg Ala Ser His Gly Thr Lys Met Met Thr Pro
 85 90 95

Glu Val Leu Ala Glu Ala Tyr Gly Lys Lys Glu Trp Lys His Phe Leu
 100 105 110

10 Ser Asp Thr Gly Met Ala Cys Arg Ser Gly Lys Tyr Tyr Phe Tyr Asp
 115 120 125

Asn Tyr Phe Asp Leu Pro Gly Ala Leu Leu Cys Ala Arg Val Val Asp
 130 135 140

15 Tyr Leu Thr Lys Leu Asn Asn Gly Gln Lys Thr Phe Asp Phe Trp Lys
 145 150 155 160

Asp Ile Val Ala Ala Ile Gln His Asn Tyr Lys Met Ser Ala Phe Lys
 20 165 170 175

Glu Asn Cys Gly Ile Tyr Phe Pro Glu Ile Lys Arg Asp Pro Gly Arg
 180 185 190

25 Tyr Leu His Ser Cys Pro Glu Ser Val Lys Lys Trp Leu Arg Gln Leu
 195 200 205

Lys Asn Ala Gly Lys Ile Leu Leu Ile Thr Ser Ser His Ser Asp
 210 215 220

30 Tyr Cys Arg Leu Leu Cys Glu Tyr Ile Leu Gly Asn Asp Phe Thr Asp
 225 230 235 240

Leu Phe Asp Ile Val Ile Thr Asn Ala Leu Lys Pro Gly Phe Phe Ser
 35 245 250 255

His Leu Pro Ser Gln Arg Pro Phe Arg Thr Leu Glu Asn Asp Glu Glu
 260 265 270

40 Gln Glu Ala Leu Pro Ser Leu Asp Lys Pro Gly Trp Tyr Ser Gln Gly
 275 280 285

Asn Ala Val His Leu Tyr Glu Leu Leu Lys Lys Met Thr Gly Lys Pro
 45 290 295 300

Glu Pro Lys Val Val Tyr Phe Gly Asp Ser Met His Ser Asp Ile Phe
 305 310 315 320

Pro Ala Arg His Tyr Ser Asn Trp Glu Thr Val Leu Ile Leu Glu Glu
 50 325 330 335

Leu Arg Gly Asp Glu Gly Thr Arg Ser Gln Arg Pro Glu Glu Ser Glu
 340 345 350

55 Pro Leu Glu Lys Lys Gly Lys Tyr Glu Gly Pro Lys Ala Lys Pro Leu
 355 360 365

Asn Thr Ser Ser Lys Lys Trp Gly Ser Phe Phe Ile Asp Ser Val Leu
 60 370 375 380

Gly Leu Glu Asn Thr Glu Asp Ser Leu Val Tyr Thr Trp Ser Cys Lys
 385 390 395 400

5 Arg Ile Ser Thr Tyr Ser Thr Ile Ala Ile Pro Ser Ile Glu Ala Ile
 405 410 415

Ala Glu Leu Pro Leu Asp Tyr Lys Phe Thr Arg Phe Ser Ser Ser Asn
 420 425 430

10 Ser Lys Thr Ala Gly Tyr Tyr Pro Asn Pro Pro Leu Val Leu Ser Ser
 435 440 445

Asp Glu Thr Leu Ile Ser Lys
 450 455

15

(2) INFORMATION FOR SEQ ID NO: 234:

20 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 234:
 Thr Ser Ser His Ser Asp Tyr Cys Arg Leu Leu Cys Glu Tyr Ile Leu
 1 5 10 15

30 Gly Asn Asp Phe Thr Asp Leu Phe Asp Ile Val
 20 25

(2) INFORMATION FOR SEQ ID NO: 235:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 327 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 235:
 Met Lys Thr Lys Asn Ile Pro Glu Ala His Gln Asp Ala Phe Lys Thr
 1 5 10 15

45 Gly Phe Ala Glu Gly Phe Leu Lys Ala Gln Ala Leu Thr Gln Lys Thr
 20 25 30

50 Asn Asp Ser Leu Arg Arg Thr Arg Leu Ile Leu Phe Val Leu Leu Leu
 35 40 45

Phe Gly Ile Tyr Gly Leu Leu Lys Asn Pro Phe Leu Ser Val Arg Phe
 50 55 60

55 Arg Thr Thr Thr Gly Leu Asp Ser Ala Val Asp Pro Val Gln Met Lys
 65 70 75 80

Asn Val Thr Phe Glu His Val Lys Gly Val Glu Glu Ala Lys Gln Glu
 85 90 95

60 Leu Gln Glu Val Val Glu Phe Leu Lys Asn Pro Gln Lys Phe Thr Ile

	100	105	110
	Leu Gly Gly Lys Leu Pro Lys Gly Ile Leu Leu Val Gly Pro Pro Gly		
	115	120	125
5	Thr Gly Lys Thr Leu Leu Ala Arg Ala Val Ala Gly Glu Ala Asp Val		
	130	135	140
	Pro Phe Tyr Tyr Ala Ser Gly Ser Glu Phe Asp Glu Met Phe Val Gly		
10	145	150	155
	Val Gly Ala Ser Arg Ile Arg Asn Leu Phe Arg Glu Ala Lys Ala Asn		
	165	170	175
15	Ala Pro Cys Val Ile Phe Ile Asp Glu Leu Asp Ser Val Gly Gly Lys		
	180	185	190
	Arg Ile Glu Ser Pro Met His Pro Tyr Ser Arg Gln Thr Ile Asn Gln		
	195	200	205
20	Leu Leu Ala Glu Met Asp Gly Phe Lys Pro Asn Glu Gly Val Ile Ile		
	210	215	220
	Ile Gly Ala Thr Asn Phe Pro Glu Ala Leu Asp Asn Ala Leu Ile Arg		
25	225	230	235
	Pro Gly Arg Phe Asp Met Gln Val Thr Val Pro Arg Pro Asp Val Lys		
	245	250	255
30	Gly Arg Thr Glu Ile Leu Lys Trp Tyr Leu Asn Lys Ile Lys Phe Asp		
	260	265	270
	Xaa Ser Val Asp Pro Glu Ile Ile Ala Arg Gly Thr Val Gly Phe Ser		
	275	280	285
35	Gly Ala Glu Leu Glu Asn Leu Val Asn Gln Ala Ala Leu Lys Ala Ala		
	290	295	300
	Val Asp Gly Lys Glu Met Val Thr Met Lys Glu Leu Gly Val Phe Gln		
40	305	310	315
	Arg Gln Asn Ser Asn Gly Ala		
	325		

45

(2) INFORMATION FOR SEQ ID NO: 236:

	(i) SEQUENCE CHARACTERISTICS:		
50	(A) LENGTH: 21 amino acids		
	(B) TYPE: amino acid		
	(D) TOPOLOGY: linear		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 236:		
55	Met Lys Thr Lys Asn Ile Pro Glu Ala His Gln Asp Ala Phe Lys Thr		
	1	5	10
	Gly Phe Ala Glu Gly		
	20		
60			

(2) INFORMATION FOR SEQ ID NO: 237:

5 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

10

Pro Val Gln Met Lys Asn Val Thr Phe Glu His Val Lys Gly Val Glu
1 5 10 15

15

Glu Ala Lys Gln Glu Leu Gln
20

(2) INFORMATION FOR SEQ ID NO: 238:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

Ser Arg Gln Thr Ile Asn Gln Leu Leu Ala Glu Met Asp Gly Phe Lys
1 5 10 15

30

Pro Asn Glu Gly Val Ile Ile
20

35

(2) INFORMATION FOR SEQ ID NO: 239:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

40

Phe Ser Gly Ala Glu Leu Glu Asn Leu Val Asn Gln Ala Ala Leu Lys
1 5 10 15

45

Ala Ala Val Asp Gly Lys Glu Met
20

50

(2) INFORMATION FOR SEQ ID NO: 240:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 192 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

60

Leu Pro Met Trp Gln Val Thr Ala Phe Leu Asp His Asn Ile Val Thr
1 5 10 15

Ala Gln Thr Thr Trp Lys Gly Leu Trp Met Ser Cys Val Val Gln Ser
 20 25 30

5 Thr Gly His Met Gln Cys Lys Val Tyr Asp Ser Val Leu Ala Leu Ser
 35 40 45

Thr Glu Val Gln Ala Ala Arg Ala Leu Thr Val Ser Ala Val Leu Leu
 50 55 60

10 Ala Phe Val Ala Leu Phe Val Thr Leu Ala Gly Ala Gln Cys Thr Thr
 65 70 75 80

15 Cys Val Ala Pro Gly Pro Ala Lys Ala Arg Val Ala Leu Thr Gly Gly
 85 90 95

Val Leu Tyr Leu Phe Cys Gly Leu Leu Ala Leu Val Pro Leu Cys Trp
 100 105 110

20 Phe Ala Asn Ile Val Val Arg Glu Phe Tyr Asp Pro Ser Val Pro Val
 115 120 125

Ser Gln Lys Tyr Glu Leu Gly Ala Xaa Leu Tyr Ile Gly Trp Ala Ala
 130 135 140

25 Thr Ala Leu Leu Met Val Gly Gly Cys Leu Leu Cys Cys Gly Ala Trp
 145 150 155 160

30 Val Cys Thr Gly Arg Pro Asp Leu Ser Phe Pro Val Lys Tyr Ser Ala
 165 170 175

Pro Arg Arg Pro Thr Ala Thr Gly Asp Tyr Asp Lys Lys Asn Tyr Val
 180 185 190

35

40 (2) INFORMATION FOR SEQ ID NO: 241:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

Leu His Tyr Phe Ala Leu Ser Phe Val Leu Ile Leu Thr Glu Ile Cys
 1 5 10 15

50 Leu Val Ser Ser Gly Met Gly Phe
 20

55 (2) INFORMATION FOR SEQ ID NO: 242:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids
 (B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

5 Gln Leu Arg Asn Gly Ile Pro Pro Gly Arg Lys Ala Leu Phe Cys Ser
1 5 10 15

Gly Lys Pro Arg Leu Phe Thr Leu Gly Gln Gly Arg Thr Cys Ala
20 25 30

10

(2) INFORMATION FOR SEQ ID NO: 243:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 39 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

20 Trp Ser Gly Leu Trp Val Thr Thr Trp Asn Gly Ser Ser Gly Glu Arg
1 5 10 15

Thr Pro Ser Pro Trp Arg Arg Lys Arg Ala Ser Gln Ser Ala Gly Arg
20 25 30

25 Ile Ala Ser Trp Met Ser Phe
35

30

(2) INFORMATION FOR SEQ ID NO: 244:

(i) SEQUENCE CHARACTERISTICS:

35 (A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

40 Glu Tyr Asn Lys Glu Ser Glu Asp Lys Tyr Val Phe Leu Val
1 5 10

45

(2) INFORMATION FOR SEQ ID NO: 245:

(i) SEQUENCE CHARACTERISTICS:

50 (A) LENGTH: 14 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

Ile Asp Val Glu Ile Ala Arg Ser Asp Cys Arg Lys Pro Leu
1 5 10

55

(2) INFORMATION FOR SEQ ID NO: 246:

(i) SEQUENCE CHARACTERISTICS:

60 (A) LENGTH: 142 amino acids

(B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

5 Met Pro Arg Cys Arg Trp Leu Ser Leu Ile Leu Leu Thr Ile Pro Leu
 1 5 10 15

Ala Leu Val Ala Arg Lys Asp Pro Lys Lys Asn Glu Thr Gly Val Leu
 20 25 30

10 Arg Lys Leu Lys Pro Val Asn Ala Ser Asn Ala Asn Val Lys Gln Cys
 35 40 45

15 Leu Trp Phe Ala Met Gln Glu Tyr Asn Lys Glu Ser Glu Asp Lys Tyr
 50 55 60

Val Phe Leu Val Val Lys Thr Leu Gln Ala Gln Leu Gln Val Thr Asn
 65 70 75 80

20 Leu Leu Glu Tyr Leu Ile Asp Val Glu Ile Ala Arg Ser Asp Cys Arg
 85 90 95

Lys Pro Leu Ser Thr Asn Glu Ile Cys Ala Ile Gln Glu Asn Ser Lys
 100 105 110

25 Leu Lys Arg Lys Leu Ser Cys Ser Phe Leu Val Gly Ala Leu Pro Trp
 115 120 125

Asn Gly Glu Phe Thr Val Met Glu Lys Lys Cys Glu Asp Ala
 30 130 135 140

(2) INFORMATION FOR SEQ ID NO: 247:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 92 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

Cys Leu Trp Phe Ala Met Gln Glu Tyr Asn Lys Glu Ser Glu Asp Lys
 1 5 10 15

45 Tyr Val Phe Leu Val Val Lys Thr Leu Gln Ala Gln Leu Gln Val Thr
 20 25 30

Asn Leu Leu Glu Tyr Leu Ile Asp Val Glu Ile Ala Arg Ser Asp Cys
 35 40 45

50 Arg Lys Pro Leu Ser Thr Asn Glu Ile Cys Ala Ile Gln Glu Asn Ser
 50 55 60

55 Lys Leu Lys Arg Lys Leu Ser Cys Ser Phe Leu Val Gly Ala Leu Pro
 65 70 75 80

Trp Asn Gly Glu Phe Thr Val Met Glu Lys Lys Cys
 85 90

(2) INFORMATION FOR SEQ ID NO: 248:

(i) SEQUENCE CHARACTERISTICS:

5

- (A) LENGTH: 123 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

10 Ala Arg Lys Asp Pro Lys Lys Asn Glu Thr Gly Val Leu Arg Lys Leu
 1 5 10 15

Lys Pro Val Asn Ala Ser Asn Ala Asn Val Lys Gln Cys Leu Trp Phe
 20 25 30

15 Ala Met Gln Glu Tyr Asn Lys Glu Ser Glu Asp Lys Tyr Val Phe Leu
 35 40 45

20 Val Val Lys Thr Leu Gln Ala Gln Leu Gln Val Thr Asn Leu Leu Glu
 50 55 60

Tyr Leu Ile Asp Val Glu Ile Ala Arg Ser Asp Cys Arg Lys Pro Leu
 65 70 75 80

25 Ser Thr Asn Glu Ile Cys Ala Ile Gln Glu Asn Ser Lys Leu Lys Arg
 85 90 95

Lys Leu Ser Cys Ser Phe Leu Val Gly Ala Leu Pro Trp Asn Gly Glu
 100 105 110

30 Phe Thr Val Met Glu Lys Lys Cys Glu Asp Ala
 115 120

35

(2) INFORMATION FOR SEQ ID NO: 249:

(i) SEQUENCE CHARACTERISTICS:

40

- (A) LENGTH: 44 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

45 Asp Ser Pro Asp Thr Glu Pro Gly Ser Ser Ala Gly Pro Thr Gln Arg
 1 5 10 15

Pro Ser Asp Asn Ser His Asn Glu His Ala Pro Ala Ser Gln Gly Leu
 20 25 30

50 Lys Ala Glu His Leu Tyr Ile Leu Ile Gly Val Ser
 35 40

55

(2) INFORMATION FOR SEQ ID NO: 250:

(i) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 101 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

His Arg Gln Asn Gln Ile Lys Gln Gly Pro Pro Arg Ser Lys Asp Glu
 1 5 10 15

5 Glu Gln Lys Pro Gln Gln Arg Pro Asp Leu Ala Val Asp Val Leu Glu
 20 25 30

Arg Thr Ala Asp Lys Ala Thr Val Asn Gly Leu Pro Glu Lys Asp Arg
 10 35 40 45

Glu Thr Asp Thr Ser Ala Leu Ala Ala Gly Ser Ser Gln Glu Val Thr
 50 55 60

15 Tyr Ala Gln Leu Asp His Trp Ala Leu Thr Gln Arg Thr Ala Arg Ala
 65 70 75 80

Val Ser Pro Gln Ser Thr Lys Pro Met Ala Glu Ser Ile Thr Tyr Ala
 20 85 90 95

Ala Val Ala Arg His
 100

25

(2) INFORMATION FOR SEQ ID NO: 251:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 115 amino acids
 30 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

Met Ser Pro His Pro Thr Ala Leu Leu Gly Leu Val Leu Cys Leu Ala
 35 1 5 10 15

Gln Thr Ile His Thr Gln Glu Glu Asp Leu Pro Arg Pro Ser Ile Ser
 20 25 30

40 Ala Glu Pro Gly Thr Val Ile Pro Leu Gly Ser His Val Thr Phe Val
 35 40 45

Cys Arg Gly Pro Val Gly Val Gln Thr Phe Arg Leu Glu Arg Glu Ser
 45 50 55 60

Arg Ser Thr Tyr Asn Asp Thr Glu Asp Val Ser Gln Ala Ser Pro Ser
 65 70 75 80

Glu Ser Glu Ala Arg Phe Arg Ile Asp Ser Val Ser Glu Gly Asn Ala
 50 85 90 95

Gly Pro Tyr Arg Cys Ile Tyr Tyr Lys Pro Pro Lys Trp Ser Glu Gln
 100 105 110

55 Ser Asp Tyr
 115

60 (2) INFORMATION FOR SEQ ID NO: 252:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

10 Thr Ala Leu Leu Gly Leu Val Leu Cys Leu Ala Gln Thr Ile His Thr
1 5 10 15
10 Gln Glu

15

(2) INFORMATION FOR SEQ ID NO: 253:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 14 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

25 Leu Pro Arg Pro Ser Ile Ser Ala Glu Pro Gly Thr Val Ile
1 5 10

30

(2) INFORMATION FOR SEQ ID NO: 254:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

Cys Arg Gly Pro Val Gly Val Gln Thr Phe Arg Leu Glu Arg Glu
1 5 10 15

40

(2) INFORMATION FOR SEQ ID NO: 255:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

50 Val Leu Glu Arg Thr Ala Asp Lys Ala Thr Val Asn Gly Leu Pro Glu
1 5 10 15

55 Lys Asp Arg Glu Thr Asp Thr Ser Ala Leu Ala Ala Gly Ser Ser
20 25 30

(2) INFORMATION FOR SEQ ID NO: 256:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 438 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

5

Met	Asn	Thr	Pro	Asn	Gly	Asn	Ser	Leu	Ser	Ala	Ala	Glu	Leu	Thr	Cys
1	5							10					15		

10

Gly	Met	Ile	Met	Cys	Leu	Ala	Arg	Gln	Ile	Pro	Gln	Ala	Thr	Ala	Ser
	20				25				30						

15

Met	Lys	Asp	Gly	Lys	Trp	Glu	Arg	Lys	Lys	Phe	Met	Gly	Thr	Glu	Leu
	35				40				45						
Asn	Gly	Lys	Thr	Leu	Gly	Ile	Leu	Gly	Arg	Ile	Gly	Arg	Glu		
	50			55			60								

20

Val	Ala	Thr	Arg	Met	Gln	Ser	Phe	Gly	Met	Lys	Thr	Ile	Gly	Tyr	Asp
	65			70			75		80						

25

Pro	Ile	Ile	Ser	Pro	Glu	Val	Ser	Ala	Ser	Phe	Gly	Val	Gln	Gln	Leu
	85			90			95								

30

Pro	Leu	Glu	Glu	Ile	Trp	Pro	Leu	Cys	Asp	Phe	Ile	Thr	Val	His	Thr
	100			105			110								

35

Pro	Leu	Leu	Pro	Ser	Thr	Thr	Gly	Leu	Leu	Asn	Asp	Asn	Thr	Phe	Ala
	115			120			125								

40

Gln	Cys	Lys	Lys	Gly	Val	Arg	Val	Val	Asn	Cys	Ala	Arg	Gly	Gly	Ile
	130			135			140								

45

Val	Asp	Glu	Gly	Ala	Leu	Leu	Arg	Ala	Leu	Gln	Ser	Gly	Gln	Cys	Ala
	145			150			155		160						

50

Gly	Ala	Ala	Leu	Asp	Val	Phe	Thr	Glu	Glu	Pro	Pro	Arg	Asp	Arg	Ala
	165			170			175								

55

Leu	Val	Asp	His	Glu	Asn	Val	Ile	Ser	Cys	Pro	His	Leu	Gly	Ala	Ser
	180			185			190								

Thr	Lys	Glu	Ala	Gln	Ser	Arg	Cys	Gly	Glu	Glu	Ile	Ala	Val	Gln	Phe
	195			200			205								

60

Val	Asp	Met	Val	Lys	Gly	Lys	Ser	Leu	Thr	Gly	Val	Val	Asn	Ala	Gln
	210			215			220								

Ala	Leu	Thr	Ser	Ala	Phe	Ser	Pro	His	Thr	Lys	Pro	Trp	Ile	Gly	Leu
	225			230			235		240						

65

Ala	Glu	Ala	Leu	Gly	Thr	Leu	Met	Arg	Ala	Trp	Ala	Gly	Ser	Pro	Lys
	245			250			255								

70

Gly	Thr	Ile	Gln	Val	Ile	Thr	Gln	Gly	Thr	Ser	Leu	Lys	Asn	Ala	Gly
	260			265			270								

Asn	Cys	Leu	Ser	Pro	Ala	Val	Ile	Val	Gly	Leu	Leu	Lys	Glu	Ala	Ser
	275			280			285								

75

Lys	Gln	Ala	Asp	Val	Asn	Leu	Val	Asn	Ala	Lys	Leu	Leu	Val	Lys	Glu
	290			295			300								

	290	295	300	
	Ala Gly Leu Asn Val Thr Thr Ser His Ser Pro Ala Ala Pro Gly Glu			
	305	310	315	320
5	Gln Gly Phe Gly Glu Cys Leu Leu Ala Val Ala Leu Ala Gly Ala Pro			
	325	330	335	
	Tyr Gln Ala Val Gly Leu Val Gln Gly Thr Thr Pro Val Leu Gln Gly			
10	340	345	350	
	Leu Asn Gly Ala Val Phe Arg Pro Glu Val Pro Leu Arg Arg Asp Leu			
	355	360	365	
15	Pro Leu Leu Leu Phe Arg Thr Gln Thr Ser Asp Pro Ala Met Leu Pro			
	370	375	380	
	Thr Met Ile Gly Leu Leu Ala Glu Ala Gly Val Arg Leu Leu Ser Tyr			
	385	390	395	400
20	Gln Thr Ser Leu Val Ser Asp Gly Glu Thr Trp His Val Met Gly Ile			
	405	410	415	
	Ser Ser Leu Leu Pro Ser Leu Glu Ala Trp Lys Gln His Val Thr Glu			
25	420	425	430	
	Ala Phe Gln Phe His Phe			
	435			
30				

(2) INFORMATION FOR SEQ ID NO: 257:

	(i) SEQUENCE CHARACTERISTICS:			
35	(A) LENGTH: 24 amino acids			
	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 257:			

40	Met Ala Phe Ala Asn Leu Arg Lys Val Leu Ile Ser Asp Ser Leu Asp			
	1	5	10	15
	Pro Cys Cys Arg Lys Ile Leu Gln			
	20			

45

(2) INFORMATION FOR SEQ ID NO: 258:

	(i) SEQUENCE CHARACTERISTICS:			
50	(A) LENGTH: 18 amino acids			
	(B) TYPE: amino acid			
	(D) TOPOLOGY: linear			
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 258:			
55	Gly Gly Leu Gln Val Val Glu Lys Gln Asn Leu Ser Lys Glu Glu Leu			
	1	5	10	15
	Ile Ala			

60

5 (2) INFORMATION FOR SEQ ID NO: 259:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

Met Cys Leu Ala Arg Gln Ile Pro Gln Ala Thr Ala Ser Met Lys Asp
1 5 10 15

15 Gly Lys Trp Glu Arg Lys Lys Phe Met Gly Thr Glu Leu
20 25

20 (2) INFORMATION FOR SEQ ID NO: 260:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

Ala Leu Thr Ser Ala Phe Ser Pro His Thr Lys Pro Trp Ile Gly Leu
1 5 10 15

30 Ala Glu Ala Leu Gly Thr Leu Met Arg Ala Trp Ala Gly
20 25

35

(2) INFORMATION FOR SEQ ID NO: 261:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

45 Glu Val Pro Leu Arg Arg Asp Leu Pro Leu Leu Leu Phe Arg Thr Gln
1 5 10 15

Thr Ser Asp Pro Ala Met Leu Pro Thr Met Ile Gly Leu Leu Ala Glu
20 25 30

50 Ala Gly Val Arg
35

55 (2) INFORMATION FOR SEQ ID NO: 262:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 109 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

5 Phe Gly Thr Arg Phe Leu Ala Asn Leu Leu Leu Glu Glu Asp Asn Lys
1 5 10 15

5 Phe Cys Ala Asp Cys Gln Ser Lys Gly Pro Arg Trp Ala Ser Trp Asn
20 25 30

10 Ile Gly Val Phe Ile Cys Ile Arg Cys Ala Xaa Ile His Arg Asn Leu
35 40 45

Gly Val His Ile Ser Arg Val Lys Ser Val Asn Leu Asp Gln Trp Thr
50 55 60

15 Gln Val Gln Ile Gln Cys Met Gln Xaa Met Gly Asn Gly Lys Ala Asn
65 70 75 80

Arg Leu Tyr Glu Ala Tyr Leu Pro Glu Thr Phe Arg Arg Pro Gln Ile
85 90 95

20 Asp Pro Ala Val Glu Gly Phe Ile Arg Asp Xaa Tyr Glu
100 105

25

(2) INFORMATION FOR SEQ ID NO: 263:

30 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

35 Glu Glu Asp Asn Lys Phe Cys Ala Asp Cys Gln Ser Lys Gly Pro Arg
1 5 10 15

Trp Ala Ser Trp Asn
20

40

(2) INFORMATION FOR SEQ ID NO: 264:

45 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

50 Gly Val Phe Ile Cys Ile Arg Cys Ala Xaa Ile His Arg Asn Leu Gly
1 5 10 15

Val His Ile Ser
20

55

(2) INFORMATION FOR SEQ ID NO: 265:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

5

Ser	Val	Asn	Leu	Asp	Gln	Trp	Thr	Gln	Val	Gln	Ile	Gln	Cys	Met	Gln
1	5								10					15	

10

Xaa	Met	Gly	Asn	Gly	Lys	Ala
	20					

15

(2) INFORMATION FOR SEQ ID NO: 266:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 245 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

Met	Asp	Leu	Leu	Gly	Leu	Asp	Ala	Pro	Val	Ala	Cys	Ser	Ile	Ala	Asn
1	5							10					15		

25

Ser	Lys	Thr	Ser	Asn	Thr	Leu	Glu	Lys	Asp	Leu	Asp	Leu	Leu	Ala	Ser
	20					25				30					

30

Val	Pro	Ser	Pro	Ser	Ser	Ser	Gly	Ser	Arg	Lys	Val	Val	Gly	Ser	Met
	35						40				45				

35	40	45
----	----	----

35

Pro	Thr	Ala	Gly	Ser	Ala	Gly	Ser	Val	Pro	Glu	Asn	Leu	Asn	Leu	Phe
50		55						60							

Pro	Glu	Pro	Gly	Ser	Lys	Ser	Glu	Glu	Ile	Gly	Lys	Lys	Gln	Leu	Ser
65		70							75				80		

Lys	Asp	Ser	Ile	Leu	Ser	Leu	Tyr	Gly	Ser	Gln	Thr	Xaa	Gln	Met	Pro
	85						90				95				

40

Thr	Gln	Ala	Met	Phe	Met	Ala	Pro	Ala	Gln	Met	Ala	Tyr	Pro	Thr	Ala
	100			105					110						

Tyr	Pro	Ser	Phe	Pro	Gly	Val	Thr	Pro	Pro	Asn	Ser	Ile	Met	Gly	Ser
	115			120				125							

45

Met	Met	Pro	Pro	Pro	Val	Gly	Met	Val	Ala	Gln	Pro	Gly	Ala	Ser	Gly
130		135						140							

50

Met	Val	Ala	Pro	Met	Ala	Met	Pro	Ala	Gly	Tyr	Met	Gly	Gly	Met	Gln
145		150							155				160		

Ala	Ser	Met	Met	Gly	Val	Pro	Asn	Gly	Met	Met	Thr	Thr	Gln	Gln	Ala
	165			170					175						

55

Gly	Tyr	Met	Ala	Gly	Met	Ala	Ala	Met	Pro	Gln	Thr	Val	Tyr	Gly	Val
	180			185				190							

Gln	Pro	Ala	Gln	Gln	Leu	Gln	Trp	Asn	Leu	Thr	Gln	Met	Thr	Gln	Gln
	195			200					205						

60

Met Ala Gly Met Asn Phe Tyr Gly Ala Asn Gly Met Met Asn Tyr Gly
 210 215 220

5 Gln Ser Met Ser Gly Gly Asn Gly Ala Ala Asn Gln Thr Leu Ser
 225 230 235 240

Pro Gln Met Trp Lys
 245

10

(2) INFORMATION FOR SEQ ID NO: 267:

(i) SEQUENCE CHARACTERISTICS:

15 (A) LENGTH: 315 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

20 Met Asp Leu Leu Gly Leu Asp Ala Pro Val Ala Cys Ser Ile Ala Asn
 1 5 10 15

Ser Lys Thr Ser Asn Thr Leu Glu Lys Asp Leu Asp Leu Leu Ala Ser
 20 25 30

25 Val Pro Ser Pro Ser Ser Gly Ser Arg Lys Val Val Gly Ser Met
 35 40 45

30 Pro Thr Ala Gly Ser Ala Gly Ser Val Pro Glu Asn Leu Asn Leu Phe
 50 55 60

Pro Glu Pro Gly Ser Lys Ser Glu Glu Ile Gly Lys Lys Gln Leu Ser
 65 70 75 80

35 Lys Asp Ser Ile Leu Ser Leu Tyr Gly Ser Gln Thr Xaa Gln Met Pro
 85 90 95

40 Thr Gln Ala Met Phe Met Ala Pro Ala Gln Met Ala Tyr Pro Thr Ala
 100 105 110

Tyr Pro Ser Phe Pro Gly Val Thr Pro Pro Asn Ser Ile Met Gly Ser
 115 120 125

45 Met Met Pro Pro Pro Val Gly Met Val Ala Gln Pro Gly Ala Ser Gly
 130 135 140

Met Val Ala Pro Met Ala Met Pro Ala Gly Tyr Met Gly Gly Met Gln
 145 150 155 160

50 Ala Ser Met Met Gly Val Pro Asn Gly Met Met Thr Thr Gln Gln Ala
 165 170 175

Gly Tyr Met Ala Gly Met Ala Ala Met Pro Gln Thr Val Tyr Gly Val
 180 185 190

55 Gln Pro Ala Gln Gln Leu Gln Trp Asn Leu Thr Gln Met Thr Gln Gln
 195 200 205

60 Met Ala Gly Met Asn Phe Tyr Gly Ala Asn Gly Met Met Asn Tyr Gly
 210 215 220

Gln Ser Met Ser Gly Gly Asn Gly Gln Ala Ala Asn Gln Thr Leu Ser
 225 230 235 240

5 Pro Gln Met Trp Lys Phe Gly Thr Arg Phe Leu Ala Asn Leu Leu Leu
 245 250 255

Glu Glu Asp Asn Lys Phe Cys Ala Asp Cys Gln Ser Lys Gly Pro Arg
 260 265 270

10 Trp Ala Ser Trp Asn Ile Gly Val Phe Ile Cys Ile Arg Cys Ala Xaa
 275 280 285

Ile His Arg Asn Leu Gly Val His Ile Ser Arg Val Lys Ser Val Asn
 15 290 295 300

Leu Asp Gln Trp Thr Gln Val Gln Ile Gln Cys
 305 310 315

20

(2) INFORMATION FOR SEQ ID NO: 268:

25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 39 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

30 Met Gln Xaa Met Gly Asn Gly Lys Ala Asn Arg Leu Tyr Glu Ala Tyr
 1 5 10 15

Leu Pro Glu Thr Phe Arg Arg Pro Gln Ile Asp Pro Ala Val Glu Gly
 20 25 30

35 Phe Ile Arg Asp Xaa Tyr Glu
 35

40

(2) INFORMATION FOR SEQ ID NO: 269:

45 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 67 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

50 Lys Tyr Gly Lys Val Gly Lys Cys Val Ile Phe Glu Ile Pro Gly Ala
 1 5 10 15

Pro Asp Asp Glu Ala Val Arg Ile Phe Leu Glu Phe Glu Arg Val Glu
 20 25 30

55 Ser Ala Ile Lys Ala Val Val Asp Leu Asn Gly Arg Tyr Phe Gly Gly
 35 40 45

Arg Val Val Lys Ala Cys Phe Tyr Asn Leu Asp Lys Phe Arg Val Leu
 50 55 60

60

Asp Leu Ala
65

5

(2) INFORMATION FOR SEQ ID NO: 270:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

15 Lys Ala Val Asp Leu Gly Arg Tyr Phe Gly Gly Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO: 271:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

30 Glu Ala Val Arg Ile Phe Phe Arg Glu
1 5

(2) INFORMATION FOR SEQ ID NO: 272:

35

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 306 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

40

Arg Met Gly Arg Phe His Arg Ile Leu Glu Pro Gly Leu Asn Ile Leu
1 5 10 15

45 Ile Pro Val Leu Asp Arg Ile Arg Tyr Val Gln Ser Leu Lys Glu Ile
20 25 30

Val Ile Asn Val Pro Glu Gln Ser Ala Val Thr Leu Asp Asn Val Thr
35 40 45

50

Leu Gln Ile Asp Gly Val Leu Tyr Leu Arg Ile Met Asp Pro Tyr Lys
50 55 60

60 Ala Ser Tyr Gly Val Glu Asp Pro Glu Tyr Ala Val Thr Gln Leu Ala
65 70 75 80

55

Gln Thr Thr Met Arg Ser Glu Leu Gly Lys Leu Ser Leu Asp Lys Val
85 90 95

60 Phe Arg Glu Arg Glu Ser Leu Asn Ala Ser Ile Val Asp Ala Ile Asn
100 105 110

Gln Ala Ala Asp Cys Trp Gly Ile Arg Cys Leu Arg Tyr Glu Ile Lys
 115 120 125

5 Asp Ile His Val Pro Pro Arg Val Lys Glu Ser Met Gln Met Gln Val
 130 135 140

Glu Ala Glu Arg Arg Lys Arg Ala Thr Val Leu Glu Ser Glu Gly Thr
 145 150 155 160

10 Arg Glu Ser Ala Ile Asn Val Ala Glu Gly Lys Lys Gln Ala Gln Ile
 165 170 175

Leu Ala Ser Glu Ala Glu Lys Ala Glu Gln Ile Asn Gln Ala Ala Gly
 180 185 190

15 Glu Ala Ser Ala Val Leu Ala Lys Ala Lys Ala Glu Ala Ile
 195 200 205

Arg Ile Leu Ala Ala Leu Thr Gln His Asn Gly Asp Ala Ala Ala
 20 210 215 220

Ser Leu Thr Val Ala Glu Gln Tyr Val Ser Ala Phe Ser Lys Leu Ala
 225 230 235 240

25 Lys Asp Ser Asn Thr Ile Leu Leu Pro Ser Asn Pro Gly Asp Val Thr
 245 250 255

Ser Met Val Ala Gln Ala Met Gly Val Tyr Gly Ala Leu Thr Lys Ala
 260 265 270

30 Pro Val Pro Gly Thr Pro Asp Ser Leu Ser Ser Gly Ser Ser Arg Asp
 275 280 285

35 Val Gln Gly Thr Asp Ala Ser Leu Asp Glu Glu Leu Asp Arg Val Lys
 290 295 300

Met Ser
 305

40

(2) INFORMATION FOR SEQ ID NO: 273:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

50 Ala Ser Tyr Gly Val Glu Asp Pro Glu Tyr Ala Val Thr Gln Leu Ala
 1 5 10 15

55 Gln Thr Thr Met Arg Ser Glu Leu Gly Lys
 20 25

(2) INFORMATION FOR SEQ ID NO: 274:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

5

Met Gln Met Gln Val Glu Ala Glu Arg Arg Lys Arg Ala Thr Val Leu
1 5 10 15

10

Glu Ser Glu Gly Thr Arg Glu Ser Ala Ile Asn
20 25

15

(2) INFORMATION FOR SEQ ID NO: 275:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

25

Leu Thr Val Ala Glu Gln Tyr Val Ser Ala Phe Ser Lys Leu Ala Lys
1 5 10 15

30

Asp Ser Asn Thr Ile Leu Leu Pro Ser Asn
20 25

35

(2) INFORMATION FOR SEQ ID NO: 276:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

40

Leu Leu Gly Ala Thr Ala Pro Leu Val Ser Leu Val Pro Glu Val Ala
1 5 10 15

45

Ala Ala Val Gly Asn Ala Gly Ala Arg Gly Ala Xaa His Trp Gly Pro
20 25 30

50

Phe Ala Glu Gly Leu Ser Thr Gly Phe Trp Pro Arg Ser Ala Arg Ala
35 40 45

55

Ser Ser Gly Leu Pro Arg Asn Thr Val Val Leu Phe Val Pro Gln Gln
50 55 60

60

Glu Ala Trp Val Val Glu
65 70

55

(2) INFORMATION FOR SEQ ID NO: 277:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 277:

Arg Met Trp Arg Asn Gly Thr His Phe Trp Glu Cys Lys Ile Val Gln
1 5 10 15

5 Pro Leu Trp Lys Thr Val Trp Trp Phe Pro Arg Lys Leu Ser Ile Glu
20 25 30

Leu Pro Glu Asn Leu Ala Ile Leu Ile Gly Thr Tyr Phe Lys
10 35 40 45

(2) INFORMATION FOR SEQ ID NO: 278:

15 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 278:

Leu Lys Arg His Phe Pro Lys Glu Ala Asn Lys His Val Lys Arg Cys
1 5 10 15

25 Ser Thr Ser Leu Asp Ile Arg Glu Ile Gln Ile Lys Ile Lys Met Arg
20 25 30

30 Tyr

30

(2) INFORMATION FOR SEQ ID NO: 279:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 328 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

Gly Thr Arg Pro Gly Glu Ser His Ala Asn Asp Leu Glu Cys Ser Gly
1 5 10 15

45 Lys Gly Lys Cys Thr Thr Lys Pro Ser Glu Ala Thr Phe Ser Cys Thr
20 25 30

Cys Glu Glu Gln Tyr Val Gly Thr Phe Cys Glu Glu Tyr Asp Ala Cys
35 40 45

50 Gln Arg Lys Pro Cys Gln Asn Asn Ala Ser Cys Ile Asp Ala Asn Glu
50 55 60

55 Lys Gln Asp Gly Ser Asn Phe Thr Cys Val Cys Leu Pro Gly Tyr Thr
65 70 75 80

Gly Glu Leu Cys Gln Ser Lys Ile Asp Tyr Cys Ile Leu Asp Pro Cys
85 90 95

60 Arg Asn Gly Ala Thr Cys Ile Ser Ser Leu Ser Gly Phe Thr Cys Gln
100 105 110

Cys Pro Glu Gly Tyr Phe Gly Ser Ala Cys Glu Glu Lys Val Asp Pro
 115 120 125
 5 Cys Ala Ser Ser Pro Cys Gln Asn Asn Gly Thr Cys Tyr Val Asp Gly
 130 135 140
 Val His Phe Thr Cys Asn Cys Ser Pro Gly Phe Thr Gly Pro Thr Cys
 145 150 155 160
 10 Ala Gln Leu Ile Asp Phe Cys Ala Leu Ser Pro Cys Ala His Gly Thr
 165 170 175
 15 Cys Arg Ser Val Gly Thr Ser Tyr Lys Cys Leu Cys Asp Pro Gly Tyr
 180 185 190
 His Gly Leu Tyr Cys Glu Glu Glu Tyr Asn Glu Cys Leu Ser Ala Pro
 195 200 205
 20 Cys Leu Asn Ala Ala Thr Cys Arg Asp Leu Val Asn Gly Tyr Glu Cys
 210 215 220
 Val Cys Leu Ala Glu Tyr Lys Gly Thr His Cys Glu Leu Tyr Lys Asp
 225 230 235 240
 25 Pro Cys Ala Asn Val Ser Cys Leu Asn Gly Ala Thr Cys Asp Ser Asp
 245 250 255
 Gly Leu Asn Gly Thr Cys Ile Cys Ala Pro Gly Phe Thr Gly Glu Glu
 30 260 265 270
 Cys Asp Ile Asp Ile Asn Glu Cys Asp Ser Asn Pro Cys His His Gly
 275 280 285
 35 Gly Ser Cys Leu Asp Gln Pro Asn Gly Tyr Asn Cys His Cys Pro His
 290 295 300
 Gly Trp Val Gly Ala Asn Cys Glu Ile His Leu Gln Trp Lys Ser Gly
 305 310 315 320
 40 His Met Ala Glu Ser Leu Thr Asn
 325
 45

(2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 amino acids
 50 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 280:
 55 Gly Lys Cys Thr Thr Lys Pro Ser Glu Ala Thr Phe Ser Cys Thr Cys
 1 5 10 15
 Glu Glu Gln Tyr Val Gly Thr Phe Cys
 20 25

(2) INFORMATION FOR SEQ ID NO: 281:

5 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

10 Cys Ala His Gly Thr Cys Arg Ser Val Gly Thr Ser Tyr Lys Cys Leu
1 5 10 15

Cys Asp Pro Gly Tyr His
20

15

(2) INFORMATION FOR SEQ ID NO: 282:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

25 Cys Ala Asn Val Ser Cys Leu Asn Gly Ala Thr Cys Asp Ser Asp Gly
1 5 10 15

30 Leu Asn Gly Thr Cys Ile Cys Ala Pro Gly Phe Thr Gly Glu Glu Cys
20 25 30

Asp

35

(2) INFORMATION FOR SEQ ID NO: 283:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 299 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

45 Met Ala Gln Asn Leu Lys Asp Leu Ala Gly Arg Leu Pro Ala Gly Pro
1 5 10 15

50 Arg Gly Met Gly Thr Ala Leu Lys Leu Leu Leu Gly Ala Gly Ala Val
20 25 30

55 Ala Tyr Gly Val Arg Glu Ser Val Phe Thr Val Glu Gly Gly His Arg
35 40 45

55 Ala Ile Phe Phe Asn Arg Ile Gly Gly Val Gln Gln Asp Thr Ile Leu
50 55 60

65 Ala Glu Gly Leu His Phe Arg Ile Pro Trp Phe Gln Tyr Pro Ile Ile
65 70 75 80

60 Tyr Asp Ile Arg Ala Arg Pro Arg Lys Ile Ser Ser Pro Thr Gly Ser

	85	90	95
	Lys Asp Leu Gln Met Val Asn Ile Ser Leu Arg Val Leu Ser Arg Pro		
	100	105	110
5	Asn Ala Gln Glu Leu Pro Ser Met Tyr Gln Arg Leu Gly Leu Asp Tyr		
	115	120	125
	Glu Glu Arg Val Leu Pro Ser Ile Val Asn Glu Val Leu Lys Ser Val		
10	130	135	140
	Val Ala Lys Phe Asn Ala Ser Gln Leu Ile Thr Gln Arg Ala Gln Val		
	145	150	155
	160		
15	Ser Leu Leu Ile Arg Arg Glu Leu Thr Glu Arg Ala Lys Asp Phe Ser		
	165	170	175
	Leu Ile Leu Asp Asp Val Ala Ile Thr Glu Leu Ser Phe Ser Arg Glu		
	180	185	190
20	Tyr Thr Ala Ala Val Glu Ala Lys Gln Val Ala Gln Gln Glu Ala Gln		
	195	200	205
	Arg Ala Gln Phe Leu Val Glu Lys Ala Lys Gln Glu Gln Arg Gln Lys		
25	210	215	220
	Ile Val Gln Ala Glu Gly Glu Ala Glu Ala Ala Lys Met Leu Gly Glu		
	225	230	235
	240		
30	Ala Leu Ser Lys Asn Pro Gly Tyr Ile Lys Leu Arg Lys Ile Arg Ala		
	245	250	255
	Ala Gln Asn Ile Ser Lys Thr Ile Ala Thr Ser Gln Asn Arg Ile Tyr		
	260	265	270
35	Leu Thr Ala Asp Asn Leu Val Leu Asn Leu Gln Asp Glu Ser Phe Thr		
	275	280	285
	Arg Gly Ser Asp Ser Leu Ile Lys Gly Lys Lys		
40	290	295	

45 (2) INFORMATION FOR SEQ ID NO: 284:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

Lys Ala Leu Ala Leu Ser Phe His Gly Trp Ser Gly Thr Gly Lys Asn			
1	5	10	15

55 Phe Val

60 (2) INFORMATION FOR SEQ ID NO: 285:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

Asn Leu Ile Asp Tyr Phe Ile Pro Phe Leu Pro Leu Glu Tyr Arg His
1 5 10 15
10 Val Arg Leu Cys Ala Arg
20

15

(2) INFORMATION FOR SEQ ID NO: 286:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

Asn Leu Ile Asp Tyr Phe Ile Pro Phe Leu Pro Leu Glu Tyr Arg His
25 1 5 10 15
Val Arg Leu Cys
20

30

(2) INFORMATION FOR SEQ ID NO: 287:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

Cys His Gln Thr Leu Phe Ile Phe Asp Glu Ala Glu Lys Leu His Pro
40 1 5 10 15
Gly Leu Leu Glu Val Leu Gly Pro His Leu
20 25
45

50

(2) INFORMATION FOR SEQ ID NO: 288:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

Pro Glu Lys Ala Leu Ala Leu Ser Phe His Gly Trp Ser Gly Thr Gly
55 1 5 10 15
Lys Asn Phe Val Ala
60 20

5 (2) INFORMATION FOR SEQ ID NO: 289:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

Asn Leu Lys Glu Lys Ile Phe Ile Ser Phe Ala Trp Leu Pro Lys Ala
1 5 10 15

15 Thr Val Gln Ala Ala Ile Gly
20

20 (2) INFORMATION FOR SEQ ID NO: 290:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

25 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

Trp Leu Pro Lys Ala Thr Val Gln Ala Ala Ile Gly Ser Val Ala Leu
1 5 10 15

30 Asp

35

(2) INFORMATION FOR SEQ ID NO: 291:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

40 His Asp Arg Thr Met Gln Asp Ile Val Tyr Lys Leu Val Pro Gly Leu
1 5 10 15

45 Gln Glu

50

(2) INFORMATION FOR SEQ ID NO: 292:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

60 Phe Ala Ser His Asp Arg Thr Met Gln Asp Ile Val Tyr Lys Leu Val

;

1 5 10 15

Pro Gly Leu Gln Glu Gly Glu
20

5

(2) INFORMATION FOR SEQ ID NO: 293:

10 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

15

Leu Val Leu Ser Leu Gly Ala Trp Gly Trp Pro Ser Thr Cys Leu Trp
1 5 10 15

20 Trp

(2) INFORMATION FOR SEQ ID NO: 294:

25

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

30 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

Gln Gly Lys Leu Gln Met Trp Val Asp Val Phe Pro Lys Ser Leu
1 5 10 15

35

(2) INFORMATION FOR SEQ ID NO: 295:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

45

Pro Pro Phe Asn Ile Thr Pro Arg Lys Ala Lys Lys Tyr Tyr Leu Arg
1 5 10 15

50

(2) INFORMATION FOR SEQ ID NO: 296:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

60

Lys Thr Asp Val His Tyr Arg Ser Leu Asp Gly Glu Gly Asn Phe Asn
1 5 10 15

Trp Arg Phe

5

(2) INFORMATION FOR SEQ ID NO: 297:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

Pro Arg Leu Ile Ile Gln Ile Trp Asp Asn Asp Lys Phe Ser Leu Asp
1 5 10 15

20

Asp Tyr Leu Gly Phe Leu Glu Leu Asp Leu
20 25

25

(2) INFORMATION FOR SEQ ID NO: 298:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

Ala Val Met Ile Gly Asp Asp Cys Arg Asp Asp Val Gly Gly Ala
1 5 10 15

35

(2) INFORMATION FOR SEQ ID NO: 299:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

45

Ile Leu Val Lys Thr Gly Lys Tyr Arg Ala Ser Asp Glu Glu Lys Ile
1 5 10 15

50

Asn

55

(2) INFORMATION FOR SEQ ID NO: 300:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 277 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

	Met	Asp	Ser	Met	Pro	Glu	Pro	Ala	Ser	Arg	Cys	Leu	Leu	Leu	Pro		
1				1		5			10					15			
5	Leu	Pro	Ala	Pro	Glu	Leu	Gly	Pro									
				20			25				30						
	Ser	Gln	Ala	Gly	Ala	Glu	Glu	Asn	Asp	Trp	Val	Arg	Leu	Pro	Ser	Lys	
10		35				40				45							
	Cys	Glu	Val	Cys	Lys	Tyr	Val	Ala	Val	Glu	Leu	Lys	Lys	Pro	Leu	Arg	
		50				55			55		60						
15	Lys	Arg	Gln	Asp	Thr	Glu	Val	Ile	Gly	Thr	Val	Tyr	Ile	Leu	Asp		
	65				70			75				80					
	Gln	Lys	Ala	Ser	Gly	Val	Lys	Tyr	Thr	Lys	Ser	Asp	Leu	Arg	Leu	Ile	
		85				90			90		95						
20	Glu	Val	Thr	Glu	Thr	Ile	Cys	Lys	Arg	Leu	Leu	Asp	Tyr	Ser	Leu	His	
	100				105				105		110						
	Lys	Glu	Arg	Thr	Gly	Ser	Xaa	Arg	Phe	Ala	Lys	Gly	Met	Ser	Glu	Thr	
25		115				120			120		125						
	Phe	Glu	Thr	Leu	His	Xaa	Leu	Val	His	Lys	Gly	Val	Lys	Val	Val	Met	
		130			135			135		140							
30	Asp	Ile	Pro	Tyr	Glu	Leu	Trp	Asn	Glu	Thr	Ser	Ala	Glu	Val	Ala	Asp	
	145				150				155			160					
	Leu	Lys	Lys	Gln	Cys	Asp	Val	Leu	Val	Glu	Glu	Phe	Glu	Glu	Val	Ile	
		165				170			170		175						
35	Glu	Asp	Trp	Tyr	Arg	Asn	His	Gln	Glu	Glu	Asp	Leu	Thr	Glu	Phe	Leu	
		180			185				185		190						
	Cys	Ala	Asn	His	Val	Leu	Lys	Gly	Lys	Asp	Thr	Ser	Cys	Leu	Ala	Glu	
40		195				200			200		205						
	Gln	Trp	Ser	Gly	Lys	Lys	Gly	Asp	Thr	Ala	Ala	Leu	Gly	Gly	Lys	Lys	
		210			215			215		220							
45	Ser	Lys	Lys	Ser	Ile	Arg	Ala	Lys	Ala	Ala	Gly	Gly	Arg	Ser	Ser		
	225			230			230		235		240						
	Ser	Ser	Lys	Gln	Arg	Lys	Glu	Leu	Gly	Gly	Leu	Glu	Gly	Asp	Pro	Ser	
		245			250			250		255							
50	Pro	Glu	Glu	Asp	Glu	Gly	Ile	Gln	Lys	Ala	Ser	Pro	Leu	Thr	His	Ser	
		260			265			265		270							
	Pro	Pro	Asp	Glu	Leu												
55			275														

(2) INFORMATION FOR SEQ ID NO: 301:

60 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 199 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

5

Met	Asp	Gly	Gln	Lys	Lys	Asn	Trp	Lys	Asp	Lys	Val	Val	Asp	Leu	Leu
1	5							10					15		

10	Tyr	Trp	Arg	Asp	Ile	Lys	Lys	Thr	Gly	Val	Val	Phe	Gly	Ala	Ser	Leu
					20			25				30				

Phe	Leu	Leu	Leu	Ser	Leu	Thr	Val	Phe	Ser	Ile	Val	Ser	Val	Thr	Ala
35						40				45					

15	Tyr	Ile	Ala	Leu	Ala	Leu	Leu	Ser	Val	Thr	Ile	Ser	Phe	Arg	Ile	Tyr
		50				55			60							

20	Lys	Gly	Val	Ile	Gln	Ala	Ile	Gln	Lys	Ser	Asp	Glu	Gly	His	Pro	Phe
	65			70			75			80						

Arg	Ala	Tyr	Leu	Glu	Ser	Glu	Val	Ala	Ile	Ser	Glu	Glu	Leu	Val	Gln
	85			90			95								

25	Lys	Tyr	Ser	Asn	Ser	Ala	Leu	Gly	His	Val	Asn	Cys	Thr	Ile	Lys	Glu
	100			105			110									

30	Leu	Arg	Arg	Leu	Phe	Leu	Val	Asp	Asp	Leu	Val	Asp	Ser	Leu	Lys	Phe
	115			120			125									

35	Ala	Val	Leu	Met	Trp	Val	Phe	Thr	Tyr	Val	Gly	Ala	Leu	Phe	Asn	Gly
	130			135			140									

40	Leu	Thr	Leu	Leu	Ile	Leu	Ala	Leu	Ile	Ser	Leu	Phe	Ser	Val	Pro	Val
	145			150			155			160						

45	Ile	Tyr	Glu	Arg	His	Gln	Ala	Gln	Ile	Asp	His	Tyr	Leu	Gly	Leu	Ala
	165			170			175									

50	Asn	Lys	Asn	Val	Lys	Asp	Ala	Met	Ala	Lys	Ile	Gln	Ala	Lys	Ile	Pro
	180			185			190									

Gly	Leu	Lys	Arg	Lys	Ala	Glu										
	195															

45

(2) INFORMATION FOR SEQ ID NO: 302:

50

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids															
(B) TYPE: amino acid															
(D) TOPOLOGY: linear															

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

55

Met	Ala	Val	Thr	Leu	Ser	Leu	Leu	Leu	Gly	Gly	Arg	Val	Cys	Ala		
1	5			10			15									

60

(2) INFORMATION FOR SEQ ID NO: 303:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 41 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 303:

Pro	Ser	Leu	Ala	Val	Gly	Ser	Arg	Pro	Gly	Gly	Trp	Arg	Ala	Gln	Ala
1															
															10
															15

Leu	Leu	Ala	Gly	Ser	Arg	Thr	Pro	Ile	Pro	Thr	Gly	Ser	Arg	Arg	Asn
															20
															25

Gly	Ser	Cys	Arg	Arg	Trp	Arg	Ala	Pro							
															30
															35

(2) INFORMATION FOR SEQ ID NO: 304:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 56 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 304:

Met	Ala	Val	Thr	Leu	Ser	Leu	Leu	Leu	Gly	Gly	Arg	Val	Cys	Ala	Pro
1															
															10
															15

Ser	Leu	Ala	Val	Gly	Ser	Arg	Pro	Gly	Gly	Trp	Arg	Ala	Gln	Ala	Leu
															20
															25

Leu	Ala	Gly	Ser	Arg	Thr	Pro	Ile	Pro	Thr	Gly	Ser	Arg	Arg	Asn	Gly
															30
															35

Ser	Cys	Arg	Arg	Trp	Arg	Ala	Pro								
															40
															45

(2) INFORMATION FOR SEQ ID NO: 305:

40

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 481 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

50

GATGTTACAC	AGCTCTTAA	TAATAGTGCC	CATAGCTGTA	ATAACAATGA	CAACAGTAGG		60
TAACGGTAGT	CATACCAACA	GTAGGGCAGT	GCATTTTATA	TTACAACCTGG	TTTCCTTGCTC		120
TAGTAGGCTT	GGGGATGGGT	GAAGACGGAC	AGGGCTGGCG	CAGACCCCTTT	CCTTCTCCTC		180
TCCAGCCCCAC	AGTGATCTGG	GCTTTTACAA	GACAGCCTGC	TTCCATTCAAG	TAGTGTTGGGA		240
AAGTTCCCTTC	TTGGCTTAGC	AAATACCCCTG	AGACCTTGTT	CAGTGGGCTG	TGTCCTCTCCC		300
60							

	TGGGATGCTG GGAGCACCAA GTGTGGCCGA GCTAGGGCTG CTGACTTCCT CTGGGCGCCT	360
	CTGGGCTGCG AGGGTCTCTT ATAGGAATTC AGGCCCTTTG CTGCTCCAAG AAATGCTGAG	420
5	GCTGTGGCA RAGGGKGTGTA CCCAAGGGGA CTCTTGCTCT GTGCTGACT TTGGGGRATC	480
	C	481
10		
	(2) INFORMATION FOR SEQ ID NO: 306:	
	(i) SEQUENCE CHARACTERISTICS:	
15	(A) LENGTH: 58 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 306:	
	CACAGCTCTT TAATAATAGT GGCCATAGCT GTAATAACAA TGACAACAGT AGGTAACG	58
25		
	(2) INFORMATION FOR SEQ ID NO: 307:	
	(i) SEQUENCE CHARACTERISTICS:	
30	(A) LENGTH: 59 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
35	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 307:	
	TGTGTCTCTC CCTGGGATGC TGGGAGCACC AAGTGTGGCC GAGCTAGGGC TGCTGACTT	59
40		
	(2) INFORMATION FOR SEQ ID NO: 308:	
	(i) SEQUENCE CHARACTERISTICS:	
45	(A) LENGTH: 85 base pairs	
	(B) TYPE: nucleic acid	
	(C) STRANDEDNESS: double	
	(D) TOPOLOGY: linear	
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 308:	
	GCGAGGGTCT CTTATAGGAA TTGAGGCCCT TTGCTGCTCC AAGAAATGCT GAGGCTGTGG	60
55	GCARAGGGKT GTACCCAAGG GGACT	85
60	(2) INFORMATION FOR SEQ ID NO: 309:	

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

Met Val Gly Pro Val Thr Leu His Lys Lys Ile His Thr Thr Thr Val
1 5 10 15

10 Leu Phe Ile Val Gln Ile His Ile Leu Leu Ile Gln Ala Ile Thr Gln
20 25 30

Ala Lys

15

(2) INFORMATION FOR SEQ ID NO: 310:

20 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 67 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

25 Leu Gln Met His Leu Met Ile Leu Gln Met Thr Gly Leu Ser Ile Leu
1 5 10 15

30 Ala Leu Leu Gly Lys Ser Thr Thr Thr Ile Val Glu Gln Lys Phe His
20 25 30

Asn Gly Lys Asn Gln Lys Ser Gly Leu Lys Glu Asn Arg Asp Lys Lys
35 40 45

35 Lys Gln Thr Arg Trp Gln Ser Thr Ala Ser Gln Lys Ile Gly Ile Thr
50 55 60

Glu Glu Arg
65

40

(2) INFORMATION FOR SEQ ID NO: 311:

45 (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 101 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

50 Met Val Gly Pro Val Thr Leu His Lys Lys Ile His Thr Thr Thr Val
1 5 10 15

55 Leu Phe Ile Val Gln Ile His Ile Leu Leu Ile Gln Ala Ile Thr Gln
20 25 30

Ala Lys Leu Gln Met His Leu Met Ile Leu Gln Met Thr Gly Leu Ser
35 40 45

60 Ile Leu Ala Leu Leu Gly Lys Ser Thr Thr Thr Ile Val Glu Gln Lys

50	55	60	
Phe His Asn Gly Lys Asn Gln Lys Ser Gly Leu Lys Glu Asn Arg Asp			
65	70	75	80
Lys Lys Lys Gln Thr Arg Trp Gln Ser Thr Ala Ser Gln Lys Ile Gly			
5	85	90	95
Ile Thr Glu Glu Arg			
10	100		

(2) INFORMATION FOR SEQ ID NO: 312:

15	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 74 amino acids		
	(B) TYPE: amino acid		
	(D) TOPOLOGY: linear		
20	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 312:		
Met Gln Thr Cys Pro Leu Val Gly Thr Leu Leu Thr Arg Asn Met Asp			
1	5	10	15
25	Gly Tyr Thr Cys Ala Val Val Thr Ser Thr Ser Phe Trp Ile Ile Ser		
	20	25	30
Ala Trp Xaa Leu Trp Lys Gly Ser Pro Ser Thr Ser Met Pro Thr Met			
30	35	40	45
Pro Glu Thr Pro Leu Arg Thr Leu Cys Cys Thr Lys Met Pro Ser Ile			
35	50	55	60
Phe Ser Ser Leu Met Thr Asp Gly Arg Ala			
40	65	70	

(2) INFORMATION FOR SEQ ID NO: 313:

45	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 78 amino acids		
	(B) TYPE: amino acid		
	(D) TOPOLOGY: linear		
50	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 313:		
Met Thr Leu Ile Gln Asn Cys Trp Tyr Ser Trp Leu Phe Phe Gly Phe			
1	5	10	15
55	Phe Phe His Phe Leu Arg Lys Ser Ile Ser Ile Phe Ser Ile Phe Leu		
	20	25	30
Val Cys Phe Arg Ile Leu Ala Leu Gly Pro Thr Cys Phe Leu Val Trp			
55	35	40	45
Phe Trp Lys Ala Phe Phe Arg His Ile Leu Ile Phe Ile Cys Leu Ser			
60	50	55	60
Arg Glu Val Phe Arg Pro Arg Cys Phe Leu Val Tyr Phe Arg			
65	65	70	75

(2) INFORMATION FOR SEQ ID NO: 314:

5

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 71 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Met Gly Thr Arg Ala Gln Val Thr Pro Gly Arg Leu Pro Ile Pro Pro
1 5 10 15

15 Pro Ala Pro Gly Leu Pro Phe Ser Ala Xaa Glu Pro Leu Gln Gly Gln
20 25 30

Leu Arg Arg Val Ser Ser Arg Gly Gly Phe Pro Gly Leu Ala Leu
35 40 45

20 Gln Leu Leu Arg Ser Glu Thr Val Lys Ala Tyr Val Asn Asn Glu Ile
50 55 60

25 Asn Ile Leu Ala Ser Phe Phe
65 70

(2) INFORMATION FOR SEQ ID NO: 315:

30

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 40 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

Met Leu Val Arg Thr Arg Pro Ser Gln Pro Leu Pro Leu Pro Gly Val
1 5 10 15

40 Gly Leu Gly Gly Pro Arg Ser Gly Asp Pro Pro Glu Ser Thr Glu Leu
20 25 30

Arg Lys Gly Pro Gly Phe Leu Ala
35 40

45

(2) INFORMATION FOR SEQ ID NO: 316:

50

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 262 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

Met Cys Pro Val Cys Gly Arg Ala Leu Ser Ser Pro Gly Ser Leu Gly
1 5 10 15

60 Arg His Leu Leu Ile His Ser Glu Asp Gln Arg Ser Asn Cys Ala Val
20 25 30

Cys Gly Ala Arg Phe Thr Ser His Ala Thr Phe Asn Ser Glu Lys Leu
 35 40 45

5 Pro Glu Val Leu Asn Met Glu Ser Leu Pro Thr Val His Asn Glu Gly
 50 55 60

Pro Ser Ser Ala Glu Gly Lys Asp Ile Ala Phe Ser Pro Pro Val Tyr
 65 70 75 80

10 Pro Ala Gly Ile Leu Leu Val Cys Asn Asn Cys Ala Ala Tyr Arg Lys
 85 90 95

15 Xaa Leu Glu Ala Gln Thr Pro Ser Val Xaa Lys Trp Ala Leu Arg Arg
 100 105 110

Gln Asn Glu Pro Leu Glu Val Arg Leu Gln Arg Leu Glu Arg Glu Arg
 115 120 125

20 Thr Ala Lys Lys Ser Arg Arg Asp Asn Glu Thr Pro Glu Glu Arg Glu
 130 135 140

Val Arg Arg Met Arg Asp Arg Glu Ala Lys Arg Leu Gln Arg Met Gln
 145 150 155 160

25 Glu Thr Asp Glu Gln Arg Ala Arg Arg Leu Gln Arg Asp Arg Glu Ala
 165 170 175

30 Met Arg Leu Lys Arg Ala Asn Glu Thr Pro Glu Lys Arg Gln Ala Arg
 180 185 190

Leu Ile Arg Glu Arg Glu Ala Lys Arg Leu Lys Arg Arg Leu Glu Lys
 195 200 205

35 Met Asp Met Met Leu Arg Ala Gln Phe Gly Gln Asp Pro Ser Ala Met
 210 215 220

Ala Ala Leu Ala Ala Glu Met Asn Phe Phe Gln Leu Pro Val Ser Gly
 225 230 235 240

40 Val Glu Leu Asp Xaa Gln Leu Leu Gly Lys Met Ala Phe Glu Glu Gln
 245 250 255

45 Asn Ser Ser Xaa Leu His
 260

50 (2) INFORMATION FOR SEQ ID NO: 317:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 190 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

Met Asp His Ser His His Met Gly Met Ser Tyr Met Asp Ser Asn Ser
 1 5 10 15

60 Thr Met Gln Pro Ser His His Pro Thr Thr Ser Ala Ser His Ser

	20	25	30
	His Gly Gly Gly Asp Ser Ser Met Met Met Met Pro Met Thr Phe Tyr		
	35	40	45
5	Phe Gly Phe Lys Asn Val Glu Leu Leu Phe Ser Gly Leu Val Ile Asn		
	50	55	60
	Thr Ala Gly Glu Met Ala Gly Ala Phe Val Ala Val Phe Leu Leu Ala		
10	65	70	75
	Met Phe Tyr Glu Gly Leu Lys Ile Ala Arg Glu Ser Leu Leu Arg Lys		
	85	90	95
15	Ser Gln Val Ser Ile Arg Tyr Asn Ser Met Pro Val Pro Gly Pro Asn		
	100	105	110
	Gly Thr Ile Leu Met Glu Thr His Lys Thr Val Gly Gln Gln Met Leu		
20	115	120	125
	Ser Phe Pro His Leu Leu Gln Thr Val Leu His Ile Ile Gln Val Val		
	130	135	140
	Ile Ser Tyr Phe Leu Met Leu Ile Phe Met Thr Tyr Asn Gly Tyr Leu		
25	145	150	155
	Cys Ile Ala Xaa Ala Ala Gly Thr Gly Tyr Phe Leu Phe Ser		
	165	170	175
30	Trp Lys Lys Ala Val Val Val Asp Ile Thr Glu His Cys His		
	180	185	190

35 (2) INFORMATION FOR SEQ ID NO: 318:

	(i) SEQUENCE CHARACTERISTICS:		
	(A) LENGTH: 123 amino acids		
	(B) TYPE: amino acid		
40	(D) TOPOLOGY: linear		
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:		
	Met Val Gln Pro Cys Gly Ala Cys Ala Lys Thr Xaa Trp Lys Ala Cys		
	1	5	10
45			15
	Ser Ser Cys Cys Ser Ser Pro Cys Cys Leu Gln Glu Arg Trp Pro Xaa		
	20	25	30
	Pro Xaa Ala Xaa Cys Pro Glu Xaa Gly Pro Ser Ser His Pro Gly Ile		
50	35	40	45
	Gln Ala Leu Cys Ala Val Ala Val Val Tyr Leu Ser Pro Ser Ser Arg		
	50	55	60
55	Leu Asp Trp Ser Leu Ala Pro Leu Phe Val Pro Ser Leu Ala Ala Gly		
	65	70	75
	Glu Thr Pro Leu Thr Gln Pro Ala Trp Ala Leu Thr Thr Asn Thr Leu		
60	85	90	95

363

Gly His Gly Gln Pro Ala Gln Asp Arg Leu Pro Ala Leu Gly His Cys
100 105 110

Ala Pro Ile Ser Val Leu Gly Leu Gly Ser Ser
5 115 120

364

Applicant's or agent's file reference number	008PCT	International application? <input checked="" type="checkbox"/> Unassigned <input type="checkbox"/>
----------------------------------------------	--------	----------------------------------------------------------------------------------------------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 75, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit April 28, 1997	Accession Number 209012
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	Lydelle Meadows Paralegal Specialist IAPD-PCT Operations (703) 305-3745
Authorized officer	

Applicant's or agent's file reference number	008PCT	International application? <input checked="" type="checkbox"/> Unsigned
----------------------------------------------	--------	-------------------------------------------------------------------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>75</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (<i>including postal code and country</i>) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit June 5, 1997	Accession Number 209089
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (<i>if the indications are not for all designated States</i>)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (<i>specify the general nature of the indications, e.g., "Accession Number of Deposit"</i>)	

For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	Lydell Meadows Paralegal Specialist IAPD-PCT Operations (703) 305-3745
Authorized officer	

Applicant's or agent's file reference number	2008PCT	International application	Unassigned
----------------------------------------------	---------	---------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>78</u> , line <u>N/A</u>		
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>		
Name of depositary institution American Type Culture Collection		
Address of depositary institution (<i>including postal code and country</i>) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America		
Date of deposit June 5, 1997	Accession Number	209090
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>		
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (<i>if the indications are not for all designated States</i>)		
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (<i>specify the general nature of the indications, e.g., "Accession Number of Deposit"</i>)		

For receiving Office use only



This sheet was received with the international application

For International Bureau use only



This sheet was received by the International Bureau on:

Authorized officer

Lydell Meadows
Paralegal Specialist
IAPD-PCT Operations
(703) 305-3745

Authorized officer

367	Applicant's or agent's file reference number	008PCT	International application?	Unassigned
-----	----------------------------------------------	--------	----------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 80, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit May 22, 1997	Accession Number 209076
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	

For receiving Office use only		For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:		
Authorized officer Lydell Meadows Paralegal Specialist IAPD-PCT Operations (703) 305-3745	Authorized officer		

368

Applicant's or agent's file reference number	008PCT	International application?	Unassigned
----------------------------------------------	--------	----------------------------	------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 82, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
<p>Name of depositary institution American Type Culture Collection</p> <p>Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America</p>	
Date of deposit May 29, 1997	Accession Number 209086
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
<p>The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")</p>	

For receiving Office use only

For International Bureau use only

<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Lydell Meadows Paralegal Specialist IAPD-PCT Operations (703) 305-3745	Authorized officer

369

Applicant's or agent's file reference number	008PCT	International application <input checked="" type="checkbox"/>	Unassigned <input type="checkbox"/>
----------------------------------------------	--------	---------------------------------------------------------------	-------------------------------------

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 83, line N/A	
B. IDENTIFICATION OF DEPOSIT <input type="checkbox"/> Further deposits are identified on an additional sheet	
Name of depository institution American Type Culture Collection	
Address of depository institution (<i>including postal code and country</i>) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit	June 19, 1997
Accession Number	209126
C. ADDITIONAL INDICATIONS <i>(leave blank if not applicable)</i> <input type="checkbox"/> This information is continued on an additional sheet	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE <i>(if the indications are not for all designated States)</i>	
E. SEPARATE FURNISHING OF INDICATIONS <i>(leave blank if not applicable)</i> The indications listed below will be submitted to the International Bureau later (<i>specify the general nature of the indications, e.g., "Accession Number of Deposit"</i>)	

For receiving Office use only



This sheet was received with the international application

For International Bureau use only



This sheet was received by the International Bureau on:

Authorized officer

Lydell Meadows
 Paralegal Specialist
 IAPD-PCT Operations
 (703) 305-3745

Authorized officer

What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
 - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
 - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
 - (f) a polynucleotide which is a variant of SEQ ID NO:X;
 - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
 - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
 - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

10

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

15

7. A recombinant vector comprising the isolated nucleic acid molecule of

claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

20

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

25 (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

30 (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in

35 ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

- (g) a variant of SEQ ID NO:Y;
- (h) an allelic variant of SEQ ID NO:Y; or
- (i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the
5 full length protein comprises sequential amino acid deletions from either the C-terminus
or the N-terminus.

13. An isolated antibody that binds specifically to the isolated polypeptide of
claim 11.

10

14. A recombinant host cell that expresses the isolated polypeptide of claim
11.

15

15. A method of making an isolated polypeptide comprising:
(a) culturing the recombinant host cell of claim 14 under conditions such that
said polypeptide is expressed; and
(b) recovering said polypeptide.

20

17. A method for preventing, treating, or ameliorating a medical condition,
comprising administering to a mammalian subject a therapeutically effective amount of
the polypeptide of claim 11 or the polynucleotide of claim 1.

25

18. A method of diagnosing a pathological condition or a susceptibility to a
pathological condition in a subject comprising:
(a) determining the presence or absence of a mutation in the polynucleotide of
claim 1; and
(b) diagnosing a pathological condition or a susceptibility to a pathological
30 condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a
pathological condition in a subject comprising:
(a) determining the presence or amount of expression of the polypeptide of
35 claim 11 in a biological sample; and
(b) diagnosing a pathological condition or a susceptibility to a pathological
condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

5 (a) contacting the polypeptide of claim 11 with a binding partner; and
(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

10 22. A method of identifying an activity in a biological assay, wherein the method comprises:

15 (a) expressing SEQ ID NO:X in a cell;
(b) isolating the supernatant;
(c) detecting an activity in a biological assay; and
(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 22.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/12125

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :435/69.1, 70.1, 71.1, 235.1, 243, 325, 410; 536/23.1, 23.5

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/69.1, 70.1, 71.1, 235.1, 243, 325, 410; 536/23.1, 23.5

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 679 016 A1 (MATSUBARA et al.) 11 February 1995, see entire document and sequence listing, especially SEQ ID NO. 12, position 585-605 versus reference sequence at position 42-62; SEQ ID NO. 13, position 1942-5189 versus reference sequence at position 1-248; SEQ ID NO. 15, position 569-817 versus reference sequence at position 1-249; SEQ ID NO. 16, position 233-586 versus reference sequence at position 1-354; and SEQ ID NO. 18, position 1309-1699 versus reference sequence at position 12-393.	1-10, 14, 15, and 21
Y	WO 96/40917 A1 (YALE UNIVERSITY.) 19 December 1996. See entire document and sequence listing, especially SEQ ID NO. 11, position 444-692 versus reference sequence at position 2-250.	1-10, 14, 15, and 21

 Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
B earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search

08 SEPTEMBER 1998

Date of mailing of the international search report

01 OCT 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

BRIAN R. STANTON

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/12125

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 95/27791 A1 (DAVIES et al.) 19 October 1995, See entire document and sequence listing, especially SEQ ID NO. 17, position 742-799 versus reference sequence at position 1334-1391.	1-10, 14, 15, and 21
Y	WO 95/14100 A1 (THE WELLCOME FOUNDATION LIMITED) 26 May 1995. See entire document and sequence listing, especially SEQ ID NO. 97, position 966-991 versus reference sequence at position 747-772.	1-10, 14, 15, 21
Y	WO 94/28133 A1 (AMGEN INC.) 08 December 1994, see entire document and sequence listing, especially SEQ ID NO. 14, position 758-808 versus reference sequence at position 1599-1649.	1-10, 14, 15, and 21
Y	WO 95/01437 A2 (REGENTS OF THE UNIVERSITY OF MINNESOTA) 12 January 1995, see entire document and sequence listing, especially SEQ ID NO. 19, position 69-122 versus reference sequence at position 604-657.	1-10, 14, 15, and 21

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/12125

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-10, 14 15 and 21

Remark on Protest

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/12125

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

C07H 21/02, 04; C12N 5/00, 5/04, 5/06, 5/10, 5/16; 15/00, 15/09, 15/10, 15/11, 15/12; C12P 21/04, 21/06

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

Databases: Genbank, embase, biosis, medline

Search Terms/Strategy: Sequence search of Sequences 11-19 and 97; ext; secret?; moore?/au; shi?/au; rosen?/au; ruben?/au; lafleur?/au; olsen?/au; ebner?/au; brewer?/au; young?/au; greene?/au; ferrie?/au; yu ?/au; ni ?/au; feng ?/au

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I:

Claims 1-10, 14, 15, and 21 drawn to a polynucleotide(s), vector(s) containing the polynucleotide, host cells containing the vector(s) which are SEQ ID NO: X or a polynucleotide encoding the polypeptide Y or a cDNA in the material deposited with American Type Culture Collection with accession number Z wherein the cDNA in Z hybridizes to X. Additionally Group I contains the first method making the cells (claim 14) containing the vector(s) containing the polynucleotide(s) and the first method of use of the cells (claim 15) to make a product. There appear to be a total of 46 polynucleotide sequences of which the first ten (10) are selected for examination and therefore, there are nine (9) remaining additional groups of four (4) polynucleotide sequences.

Group II:

Claims 11, 12, 16, and 23 drawn to polypeptides and/or fragments thereof with the amino acid sequence defined by SEQ ID NO: Y as found in the material deposited with the American Type Culture Collection with accession number Z. There appear to be a total of 74 polypeptide sequences and therefore 73 additional species of proteins.

Group III:

Claim 13, drawn to an antibody that binds to a polypeptide with the amino acid sequence defined by SEQ ID NO: Y as found in the material deposited with the American Type Culture Collection with accession number Z. There appear to be a total of 74 antibodies that correspond to the SEQ ID NOs: for the "Y" and "Z" sequences and therefore 73 additional species of proteins.

Group IV:

Claim 17, drawn to a process of preventing, treating, or ameliorating a medical condition by administering a polypeptide or a polynucleotide which a second/alternative process of use of the second product and of an alternative process of use of the first claimed product in Group I.

In Group IV, and where additional fees are paid, the claims are searched only insofar as they are applicable to the selected polypeptide and its corresponding SEQ ID NO: as the first species as directed to a process practiced using a polypeptide. The second species is the practice of the process using a polynucleotide. In each instance, the same selected polypeptide as for the first species of Group II and for the first 10 polynucleotide sequences for Group I would be examined. Applicant may elect to pay additional fees for each additional of the 73 different polypeptide species beyond the first one (1) polypeptide and/or the first 10 polynucleotides as set forth in the above paragraphs directed to Group I and II.

Group V:

Claim 18, drawn to a method of diagnosis of a pathological condition an another alternative process of use of the first claimed product in Group I. Additionally Group V contains indica that there are a total of 46 polynucleotide sequences and therefore, nine(9) additional groups of four (4) polynucleotide sequences beyond the first ten (10) sequences.

Group VI:

Claim 19, drawn to a method of diagnosis of a pathological condition an another alternative process of use of the polypeptide. There appear to be a total of 74 polypeptide sequences and therefore 73 additional species of proteins.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/12125

Group VII:

Claim 20, drawn to a method of identification of a binding partner for a polypeptide. There appear to be a total of 74 polypeptide sequences and therefore 73 additional species of proteins.

Group VIII:

Claim 22, drawn to a method of identification of function of a protein is another alternative process of use of the product in Group I. Additionally Group V contains indica that there are a total of 46 polynucleotide sequences and therefore, nine(9) additional groups of four (4) polynucleotide sequences beyond the first ten (10) sequences.

The inventions listed as Groups I through VIII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

Claims of Group I are drawn to nucleotides, nucleotide constructs, and/or methods requiring the use of nucleotides or nucleotide constructs that contain more than ten individual, independent, and distinct nucleotide sequences in alternative form. Accordingly, these claims are subject to lack of unity as outlined in 1192 O.G. 68 (19 November 1996).

For Group I, the first ten (10) of the individual polynucleotide sequences designated as "X" by SEQ ID NO: as set forth in the application (see for example page 29+ and/or the SEQUENCE LISTING) are included for search. The corresponding SEQ ID NO: for "Y" and "Z" for each selected "X" should also be noted. The search of the no more than ten sequences may include the complements of the selected sequences and, where appropriate, may include subsequences within the selected sequences (e.g., oligomeric probes and/or primers).

In Group IV (as directed to the species which are polynucleotides) should applicant pay the additional fee for the second appearing species in Group IV which are polynucleotides, first ten (10) of the individual polynucleotide sequences designated as "X" by SEQ ID NO: as set forth in the application (see for example page 29+ and/or the SEQUENCE LISTING) are included for search of Group IV should the fees for Group IV be paid. This is also applied to Groups V and VIII. The corresponding SEQ ID NO: for "Y" and "Z" for each selected "X" should also be noted. The search of the no more than ten sequences may include the complements of the selected sequences and, where appropriate, may include subsequences within the selected sequences (e.g., oligomeric probes and/or primers).

Where Applicant may elect to pay additional fees for a search of sequences beyond the initial ten (10) polynucleotide sequences, and in accordance with 1192 O.G. 68 (19 November 1996), applicant may select additional groups of polynucleotides consisting of four (4) sequences beyond the initial ten (10) sequences for Group I which would then be searched with Group I upon payment of the requisite fees for the requisite Groups beyond Group I.

As to the polypeptides of Groups II, III, IV (as directed to a species which is a polypeptide), VI, and VII each is a distinct and different protein. Should additional fees for the above indicated Groups be paid, the first amino acid sequence identified from the SEQUENCE LISTING by applicant would be searched with the additional group for which the additional search fees were paid.

Applicant may select additional proteins and/or antibodies to be searched by specifying the appropriate SEQ ID NOs and payment of the requisite additional fees for each single additional particular species that are selected beyond the one (1) protein identified by SEQ ID NO:.

The SEQ ID NOs in Group I define, absent evidence to the contrary, structurally distinct and different proteins. Note the present application written description (page 5+) refers to the protein encoded by gene 1 as likely to be involved in promotion of a variety of cancers whereas gene 2 (pages 6-7) is directed to apparently a variety but not correlated immune system disorder(s) whereas gene 3 (pages 7-8) is asserted at page 7 to be a mediator of ligand dependent AF-2. Each of which and absent factual evidence to the contrary, are directed to genes encoding distinct and different proteins and are therefore distinct and different genes and appear to map to different chromosomes.

As to the protein of Group II and the antibody of Group III, each is distinct and different for the reasons indicated in the preceding paragraph and because the proteins have distinct and different chemical, physical, and biological properties from that of DNA/polynucleotides/vectors and cells containing same.

Groups IV through VIII are directed to alternative processes of use of the Group I and II compositions where